DOCUMENT RESUME

ED 223 669 TM 820 794

AUTHOR	Peck, Hugh I.; And Others
TITLE	Papers to Accompany The Symposium "Using Statewide Basic Skills Tests to Make Promotion Decisions: Political and Psychometric Issues."
SPONS AGENCY	Louisiana State Dept. of Education, Baton Rouge.
PUB DATE	Mar 82
NOTE	103p.; Papers presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, March 20-22, 1982).
PUB TYPE	Collected Works - Conference Proceedings (021) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Basic Skills; Compensatory Education; Elementary School Students; Elementary Secondary Education; Evaluation Methods; Item Banks; Latent Trait Theory; *Minimum Competencies; *Minimum Competency Testing; School Districts; Secondary School Students; Special Education; State Legislation; *State Programs; State School District Relationship; *Student Promotion; Testing Programs
IDENTIFIERS	*Louisiana; Louisiana State Department of Public Education

ABSTRACT

These papers were presented at the symposium "Using Statewide Basic Skills Tests to Make Promotion Decisions: Political and Psychometric Issues." The development of basic skills testing and the legislative mandate that needed to be implemented for minimum competency testing in Louisiana are discussed by Hugh I. Peck. Rebecca S. Christian examines the involvement of local educators in the development of a basic skills testing program. A statewide planning commission assured involvement of all local education agencies. Ellen B. Gillespie considers the implications of the testing program for special education students. Federal legislation assured the provision of educational and related services to handicapped children. Special education students who were not addressing state minimum skills were in an "alternative to regular placement." They were not included in the Basic Skills Testing Program. An individual educational prescription (IEP) was developed for each of these students. Policies are being developed to assure a test which is fair for all students. The development and use of an item bank for the state testing program is described by Craig N. Mills. Ellen Pechman discusses the effects of promotional testing on a large city school system. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED223669

Papers to Accompany
The Symposium
"Using Statewide Basic Skills Tests to Make Promotion Decisions: Political and Psychometric Issues"

Tm 820 794

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Williams, Jr.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This public document was published at a cost of $3.50 per copy by the Louisiana Department of Education, Bureau of Accountability, P.O. Box 44064, Baton Rouge, LA 70804, pursuant to LA. R.S. 17:24.4. This material was printed in accordance with the standards for printing by state agencies established pursuant to R.S. 43:31. Printing of this material was purchased in accordance with the provisions of Title 43 of the Louisiana Revised Statutes.

THE ROLE OF BASIC SKILLS TESTING IN LOUISIANA

Hugh I. Peck
Louisiana Department of Education

Introduction

It may seem redundant to write another paper or organize another symposium on minimum competency testing. Certainly, the issues surrounding MCT have been extensively explored, via professional journals and public media as well. I doubt if more than 20 hours of testimony have been heard on any other single issue in education in the manner of the MCT "trial" which delved into the pros and cons of this issue. Certainly, those who were recognized as experts in the field had a tremendous opportunity to get their views before the public in a unique way. The picture, however, must be viewed from another angle by the Louisiana Department of Education, where we have not only the joy and pleasure of entering into the debate, but also the reality of implementing a unique legislative mandate for MCT.

Politically--How We Got Where We Are.

First, however, it is necessary to review some of the politics and sociology of how such laws come about. The drive for public education to be accountable did not begin in Louisiana; nevertheless, the speed and extensiveness with which such a law was passed in Louisiana leave one with the inescapable conclusion that the voters of our state wanted legislation that assured accountability in our educational community. There is no doubt that the Louisiana Legislature, through the State Board and the Department, wanted to get education on the move in our state. The issues of graduates who cannot read, write or compute, of social promotion, of teachers who cannot or will not teach, and school

administrators who cannot or will not administer were raised in our state as they were in others. Our public needed renewed faith in their public schools.

Prior to 1976, there was little or no curriculum direction from the Department to our LEAs. There were few, if any, educational standards or programs of quality control. The public and the Legislature wanted, and Superintendent J. Kelly Nix provided, the leadership to assure at least minimum educational standards were established throughout our state.

The response to the public's demand for quality education and quality assurance was the enactment of 15 major pieces of legislation in a six-year period. Thus, the Superintendent, the Board and the Department will be totally involved in its implementation over the next 10 years. (Figure 1 is a synopsis of the legislation affecting elementary and secondary education.) For the record, legislation relating to higher education, especially teacher preparation, was nearly as extensive but is not germane to the MCT issue.

Minimum Competency Testing--The Most Important Legislation

Three of these pieces of legislation are the foundation for our consideration of the psychometric and politic issues of minimum competency testing--the School Accountability Act (1977), the Competency Based Education Act (1979), and the Compensatory/Remedial Education Act (1980).

The accountability law directed the Department to establish minimum standards in the basic skills areas, a statewide assessment program, a program for personnel evaluation for employees of the SEA, and the LEAs (a law requiring monitoring of this program was enacted

FIGURE 1

SYNOPSIS OF LEGISLATION AFFECTING ELEMENTARY AND SECONDARY EDUCATION

STUDENT IMPROVEMENT

Minimum Competency Standards
Act No. 621 (R.S. 17:391.1-17:391.10)

As a first step to improvement of student performance, minimum competency standards were established to define skill areas in reading, writing and mathematics for grades K-12. Teams of teachers, principals and supervisors from across the state worked with State Department of Education personnel and university personnel to establish this all-important first step.

Curriculum Guides
Act No. 750 (R.S. 17:24.4)

Curriculum guides, based upon the skill areas defined in minimum competency standards, have been developed for all required subjects. To accomplish this task, more than 600 teachers, supervisors and university personnel contributed their skills. The curricula materials were field-tested in 44 school systems during the 1980-81 school year and will be used statewide in the 1981-82 school year.

State Assessment
Act No. 621 (R.S. 17:391.1-17:391.10)

Criterion-referenced tests, based upon skills identified in the minimum competency standards, were piloted and administered to grades 4, 8 and 11 in reading during the fall of 1977-78. During the 1978-79 school year tests were developed in mathematics, and in 1979-80 they were completed for writing. Now, educators and parents had one more tool to use in determining how well students were learning essential skills. For the first time in many years, we have a "match" between skills (minimum competency standards), teaching (curriculum guides) and testing programs.

In 1981, the State Assessment Program was modified to a grade 3, 7 and 10 spring testing cycle in order to allow teachers to have access to student's test results at an earlier time. This will aid the teacher by allowing more time for diagnostic prescriptive teaching.

Basic Skills Testing
Act No. 750 (R.S. 17:24.4)

Beginning in the 1980-81 school year, a test was developed and piloted in basic skills for grade 2. Again, this test was based on the minimum competency standards in reading, writing and mathematics. In the spring of the 1981-82 school year this test will be used as the principal basis for a student's promotion to 3rd grade. Each year after the 1981-82 school year, we will add a grade level to the basic skills testing program, phasing out the 3, 7 and 10 program, until all grade levels are being tested each year.

Pupil Progression Plans
Act No. 750 (R.S. 17:24.4)

Each local school district has developed a Pupil Progression Plan which defines promotion standards from one grade level to the next.

Compensatory Education
Act No. 433 (R.S. 17:39.4-17:400)

The next logical question is what to do about students who do not achieve up to expectations. The answer has been the development of plans for compensatory (or remedial) education. Specifically, funds were used to support pilot, planning or demonstration projects in parish or city school systems. These projects will provide models for other systems to emulate in setting up their own compensatory programs.

Coordination and Involvement of All Federal Instructional Programs
Act No. 750 (R.S. 17:24.4)

Emphasis is being placed upon coordination of regular instructional programs with Title I, Title IV, Vocational Education and Special Education programs in order to provide a more effective total instructional program for all children.

An example of the way this coordination is being accomplished is through the PAAS Program, which was mandated in Act 621.

PERSONNEL IMPROVEMENT

Job Descriptions and Personnel Evaluation
Act No. 621 (R.S. 17:391.1-17:391.10)

Job descriptions have been written by each local school district and the State Department for all certificated personnel. Additionally, a uniform assessment program for all educational employees, including State Department of Education personnel and local school system employees, has been implemented.

Uniform Certification Standards
Act No. 16 (R.S. 17:7(6)(b).)

On or after September 15, 1978, any person applying for initial certification must pass a proficiency examination. After validation by 522 subject matter experts, the National Teachers Examination was selected for use in Louisiana.

Professional Practices Commission
Act No. 236 (R.S. 17:1331-17:1342)

The duty of this Commission is to investigate any reasonable alleged grievances. While the Commission does not have direct authority over the issuance or suspension of licenses, it can recommend to the BESE that such action be taken. The Commission shall be composed of members who are certified to teach.

Professional Improvement Program
Act No. 207 (R.S. 17:3601-17:3661)

This program provides the mechanism by which educational employees in public elementary and secondary schools will have an opportunity for professional growth and improvement and, at the same time, enjoy significant salary increases.

Tuition Exemption Program for Teachers
Act No. 20 (R.S. 17:7.3)

Louisiana teachers may now attend public state colleges and universities on a tuition-exemption basis. Acts 718, 261 and 379 increased the scope of personnel who are eligible.

Paid In-Service Days
Act No. 10 (R.S. 17:24.1)

This program provides 2 days of salary-paid in-service training for educators. The training is based upon local needs and designed by teachers and principals on the district level.

Teacher Education Programs
Act No. 756 (R.S. 17:7.1)

In 1977 a law was passed to require a 2.2 grade point average for entrance and a 2.5 grade point average for completion of a teacher preparation program. Additionally, the law requires 6 hours of reading instruction for secondary majors and 9 hours for elementary majors. Student teaching time was increased from 90 clock hours to 270 clock hours.

Act No. 757 (R.S. 17:7.2)

Practical experiences in classroom situations are now required for students in teacher preparation programs. "In-school" experiences must begin at the sophomore level and include schools of varied socio-economic and cultural characteristics.

Accreditation (PAAS)—Program Assistance and Approval System

Act No. 621 (R.S. 17:391.1-17:391.10)

The goal of PAAS is to coordinate all monitoring activities into one on-site visitation that includes compliance activities of special education, vocational education, and Titles I and IV programs.

PAAS standards focus on aspects of both school and personnel improvement. These standards include being certain that districts are fully implementing the following:

1. Minimum Skills/Maximum Goals Standards
2. Curriculum Guides
3. Pupil Progression Plans
4. Personnel Evaluation Plans

the following year), a study of school efficiency and effectiveness and a program of school, program, and LEA accreditation. After reviewing many other state accountability plans, I am convinced that this is the most ambitious and extensive program of accountability that any state has undertaken.

The competency based education law passed in 1980 directed the Department to:

- coordinate all curriculum efforts state and federal, including regular, vocational and special education,
- develop curriculum guides in all required subjects,
- develop and establish a basic skills testing program beginning in grade two, adding a grade each year through grade twelve in 1992,
- direct and require of each LEA a pupil progression plan that assures mastery of minimum skills before promotion,
- provide remediation for students, and
- coordinate the programs of teacher preparation on a statewide basis.

The compensatory/remedial education law directs the state to provide remediation for those students who do not reach mastery on the minimum standards, to establish and evaluate 16 pilot programs in remediation (funded only one year) and, strangely enough, each year, to close down the five least effective LEA remediation programs. Those programs would be required to re-gear before being re-funded. The regular school year programs were to be funded through the minimum foundation formula, and the summer school through per-pupil allocations. Unfortunately, the impact of this legislation was weakened when funds for a second year of the remedial pilot projects were not forthcoming, and six to eight valuable demonstration centers to be selected from the 16 pilots could not be established. The rhythm of experiment, pilot, demonstrate, and implement was rudely interrupted.

FIGURE 2

MATHEMATICS

* Introduced
** Ongoing
*** Mastery

MINIMUM STANDARDS

WHOLE NUMBER OPERATIONS continued	GRADES										
	Read-iness	1	2	3	4	5	6	7	8	9	10
25. Add two four-digit numbers. (regrouping)					*	***					
26. Add two five-digit numbers. (regrouping)					*	***					
27. Add columns of numbers with five or less digits, five or less addends.							*	***			
28. Estimate a sum to the appropriate place value.						*	**	**	**	***	
29. Identify the parts of an addition problem.		*	**	***							
30. Demonsirate a knowledge of subtraction facts with minuends to ten.		*	***								
31. Subtract a one-digit number from a two-digit number (minuends to 18) using concrete objects.		*	**	***							
32. Subtract a one-digit number from a two-digit number (minuends to 18) using basic facts.		*	**	***							

Mandates for Change

The Department has the responsibility to respond to laws affecting education. This paper will specifically address legislation that affects the basic skills testing program.

Minimum Competency Standards

First, the Department, using extensive local educator participation, developed statements of minimum standards in reading, mathematics, and writing in grades K-12. Later, a statement of twelfth grade minimum competencies was established. Another paper will chronicle the development of those standards. The standards established at what level a skill was to be introduced, reinforced, and mastered within the K-12 program. Figure 2 illustrates the format of those standards.

State Assessment Program

Louisiana has two separate testing programs mandated by different pieces of legislation. The State Assessment Program was developed first with tests in reading (1978), mathematics (1979), and writing (1980) being administered in grades 4, 8 and 11 through the 1981-82 school year. Items for this fall assessment program were developed directly from the mastery level of the previous year's skill statements. Because of a need to provide more time for diagnostic-prescriptive teaching, the decision was made to move the State Assessment from a fall cycle to a spring testing cycle beginning in 1981. Another important reason for the move from fall to spring testing was to more easily phase in our second testing program, the Basic Skills Test.

Experience Counts

It is necessary to point out that prior to entering into our second legislated test, the Basic Skills Testing program, the Department already had extensive testing experience. Not only had we learned a great deal about criterion-referenced test development based upon Louisiana competencies, but also we had gained valuable insight into the logistics of large-scale testing. We had learned how to successfully bring together certified, experienced teachers and Department staff, together with ETS consultants, in a manner that provided quality control over our item writing efforts. (Another paper in this symposium will chronicle our item development process.) Further, we had learned how to manage enormous amounts of paperwork and how to develop an efficient reporting system.

Basic Skills Testing

Armed with these skills, we developed a Basic Skills Testing program by writing items in language arts and mathematics, based upon the second grade competency standards. The test was piloted in the spring of 1981 and will be given to all second graders in our state in the spring of 1982. The second grade Basic Skills Test will become the principal criterion for promotion. Naturally, this test has aroused a great deal of interest among parents and educators.

Curriculum Guides

The next logical step in creating a "match" between our skills standards and our testing programs was the creation of curriculum guides in required subject areas. The development, piloting, evaluation, and implementation of these guides is another chronicle related to our basic skills testing program, but not necessary to this

series. The purpose was to assure that the instructional program was in place before the basic skills testing began. Such an effort is not only sound educationally; it is a legal necessity. Another major purpose was to assure that minimum standards did not become maximum goals. Figure 3 is a page illustrating the format and content of those guides. The synchronization of standards, curriculum, and testing is basic to our program.

Pupil Progression Plans

The development of pupil progression plans for each parish really started to cast the anticipated light on the minimum competency testing effort (later renamed Basic Skills Testing). The State Board plans, through guidelines from the Department, directed each LEA to establish written policy relating to the promotion decision processes. The law assured that the LEA guidelines required extensive input from the educational and parental community of each parish. It further stated that, though other factors were to be considered, the results of the student's performance on the basic skills test were to be the principal criteria in the promotion decision. The plan was to further detail the LEA plan for remediation. Figure 4 is a table of contents illustrative of the extensive pupil progression plans. For more than a year, pupil progression plans became the hottest educational issue in Louisiana. Painfully, perhaps, our LEAs used these plans to develop the broad educational programs necessary to face the MCT. When the pupil progression plans were taken before the State Board for approval, the meeting was long, heated, and had extensive citizen input.

FIGURE 3

COURSE CONTENT

Second Grade

Curriculum Outline	Performance Objectives	Activity Page Number
II. Numeration		
A. Place Value	A. To develop an understanding of place value, the student will be able to:	
	***1. Group concrete and semi-concrete objects in tens.	291
	***2. Recognize place value for ones, tens.	297
	*3. Recognize place value through hundreds.	287
	*4. Recognize place value through thousands.	287
B. Counting and Writing	B. To exhibit proficiency in counting and writing numbers, the student will be able to:	
	***1. Count to one hundred by ones, twos, fives and tens.	250
	***2. Write to one hundred by ones, fives and tens.	271
	**3. Write to one hundred by twos.	271
	*4. Read and write numbers through one thousand.	282

23

CONTENT: Numeration

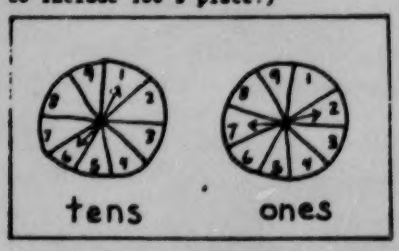
OBJECTIVE: The student will be able to recognize place value:

(1)	ones and tens	1-2	*2
(2)	through hundreds	2-3	*3
(3)	through thousands	3-4	*4
(4)	through hundred thousands	4-5	*5

ACTIVITIES

1. Play a card game with two stacks of small cards, expanded numerals for 10-30 in one stack and the two-digit numerals 10-30 in another stack. Each player may be given several expanded numeral cards and in turn draw a two-digit numeral. Each time the player gets a match, the player lays down the two matching cards. The player to discard all the cards first wins. If both children have unmatched cards, shuffle the discard pile, and reuse. Cards can be made in attractive shapes (fish, flowers, etc.).

2. See diagram below. Spin both spinners and have the children write on scratch paper the two-digit numeral for the number shown by the spinners. Have a child write the two-digit numerals on the board and have another child read the numeral aloud and tell how many tens and how many ones. (Can adapt to three spinners to include 100's place.)



3. "I Am Thinking of a Number" may be played with tens and ones. The child who is "it" might say, "I am thinking of 2 tens plus 4 ones." Another child writes or shows the numeral card for the two-digit numeral or 24 in this case. If the child is correct, that child becomes "it."

FIGURE 4

PUPIL PROGRESSION PLANS

TABLE OF CONTENTS

PREFACE 3.0

SECTION I. GENERAL PROCEDURES 4.0

A. Development of the Plan 4.0

1. Committee of Educators 4.0

2. Committee of Parents 5.0

B. Public Notice 6.0

C. School Board Policies Relative to Pupil Progression 7.0

SECTION II. PLACEMENT POLICIES 8.0

A. Regular Placement 8.0

1. Promotion: Grades K-12 . . . . . 8.0

2. Retention: Grades K-12 . . . . . 9.0

3. Compensatory Programs and/or Remediation: Grades K-12 . . . . . 10.0

a. Existing Compensatory and/or Remedial Programs 10.0

b. Projected Compensatory/Remedial Program for Second Graders 10.2

c. Projected Program for Remedial Learners (1-12). 10.1

4. Exceptional Students: Specially Designed Regular Instructional Program 11.0

5. Acceleration 12.0

6. Transfer Students 13.0

B. Alternatives to Regular Placement 14.0

1. Adult Education 14.0

2. Early College Admission 14.3

3. Exceptional Students: Separate Minimum Standards 14.4

4. Foreign Language Students 14.5

C. Review of Placement 15.0

D. Policies on Records and Reports 16.0

E. Policies on Due Process 17.0

APPENDIX A. DEFINITION OF TERMS 18.0

A. State Terms 18.0

B. Local Terms 18.1

APPENDIX B. LOUISIANA LITERACY ASSESSMENT PROGRAM 18.3

APPENDIX C. LOCAL CRITERIA 18.4

Basic Skills Programs: Grades 1-8 . . . . . 18.4

Local Testing Program 18.6

Minimum Reading Requirements 18.8

Primary Basic Skills Program 18.9

Projected Intermediate Basic Skills Program 18.15

Projected Junior High Basic Skills Program 18.21

APPENDIX D. GRADING GUIDELINES. 18.25

Academic Grades 18.25

Conduct Grades 18.28

CHECKLIST. 18.31

Compensatory/Remedial Education

Since the law requires compensatory/remedial education, extensive plans for those programs are under way. While remedial education is not crucial to the exploration of basic skills testing issue, it certainly is an important educational outcome that can't be ignored. As an indication that we're taking the problem of remediation seriously, it should be noted that 2.1 million dollars of state funds have been allocated for the 1982 summer program. Four and one-half million for the 1982-83 school year program and 2.4 million for the 1983 summer program at this point seem assured. Louisiana is extremely cognizant of its responsibility to student remedial and compensatory education. The chief state school officer has repeatedly stated that he will no longer support the program if remedial funds are not forthcoming in legislative budgets.

Parental Involvement

A priority issue in the Department has been parental awareness and involvement in basic skills testing. A series of television programs and extensive newspaper and radio coverage have been part of this campaign. Department publications explaining the program have been numerous, and illustrations are attached. By far the most effective and successful program of parental involvement has been our Calendars of Skills, presently distributed to all families with students in grades one, two, or three. Figure 5 is a page illustrative of those calendars. As you can see, a daily skill development activity is suggested for the student. Though not original with Louisiana, these were an overwhelming success. They're our free "best sellers." The Department has printed and distributed at least a quarter of a million of these.

FIGURE 5

CALENDARS OF SKILLS

FEBRUARY GRADE 2 1982

SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
	1 Add Have your child solve this number puzzle □ = 3 ☆ = 4 ★ = 7 □ ★ ☆ + □	**2** Have your child guess whether or not the ground hog will see his shadow GROUND HOG DAY	**3** Main Idea Have your child write these words in a way that helps show their meaning: tall, fat, scary, wavy, small WAVY	**4** Sequence Have your child act out in order what he or she does to get ready for school in the morning (get out of bed, get dressed, etc.)	**5** Place Value Have your child copy 4 numbers from grocery ads and write them as below (25 = 20 + 5)	**6** On a trip to the grocery store, have your child point out 3 the things that are red.
7 Have your child make his or her own mailbox using a shoebox and decorate for Valentine's Day	**8** Word Meanings From newspapers and magazines, have your child cut words and pictures that have special meaning during February (heart, George Washington)	**9** Calendar Have your child find and circle these dates on the calendar Valentine's Day (Feb 14), Washington's Birthday (Feb 22), Lincoln's Birthday (Feb 12)	**10** Cut out 10 heart shapes. Write a word from the Harris-Jacobsen List" on each heart shape. Have your child practice the words.	**11** Subtract Have your child subtract the date of Valentine's Day from the number of days in February	**12** Have your child add the dates of Washington's (Feb 22) and Lincoln's (Feb 12) birthdays LINCOLN'S BIRTHDAY	**13** Read a story to your child and discuss with him or her whether or not the story could have really happened.
14 VALENTINE'S DAY 	**15** Time Fold a piece of paper into 3 sections. Let your child label the sections "morning," "noon" and "night." Then, have him list activities he does [illegible]	**16** Write Numerals Have your child write the numerals 1-100 around the edge of a sheet of newspaper	**17** Problem Solving Have your child solve this problem: There are 54 carrots in the garden. The Carrot-Grabber ate 2. How many are left?	**18** Time Have your child look in the TV section of the newspaper and find a show that starts at 8:00	**19** Help your child do number sentences by using objects. Ex. 2 balls + 3 balls = 5 balls. Let the child rearrange the balls to make other number sentences.	**20** Make Smiley Pancakes for Saturday breakfast. Have your child help by making eyes and mouth with butter
21 Number Series Read this number series to your child and have him or her clap when you say a number that does not belong: 23, 24, 25, 29, 26	**22** WASHINGTON'S BIRTHDAY Call out the ages of family members and have your child write the numerals for each	**23** Consonants Have your child clap every time he or she hears a word that begins with /s/. Say Wacky Willie went to the window to watch Wanda with her wagon	**24** Vowels Label 2 cups, 1 with "long i," 1 with "long e." Have your child draw a picture for each of these words: bike, five, feet, sheet. Have your child put each picture in the correct cup	**25** Symbols Ask your child to write for you the symbol for addition (+), subtraction (-), and equals to (=). Have him show you he knows how to use these symbols in several math problems.	**26** Help your child do the following addition problem 57 + 32	**27** Have your child compare himself or herself to a good friend and tell how they are alike and how they are different.
28 Ask your child what job he or she thinks he or she would like to have sometime. Have your child draw/color a person working on that job	SHARE YOUR TIME with YOUR CHILD. 					

* PROVIDED AT THE BACK OF THE CALENDAR

NOTE: The Louisiana State Dept. of Ed. [illegible] acknowledges the contribution of the Duval County Florida Schools for selected materials in this calendar

A Conclusion Becomes the Introduction

All this leads up to the major issue of these related papers--the basic skills testing program in Louisiana. The sequence of this development is present in Figure 6. Note that as Basic Skills Tests are phased in, State Assessment is shown as phased out. More recent decisions have been reached that will move state assessment into the non-basic skills areas, such as science and social studies, rather than phase out the program. This change has long been the desire of the curriculum specialists in these areas.

No issue now preys stronger on our minds than how we keep the minimum standards from becoming the maximum goals, or the total program. Repeatedly, LEAs are encouraged to strive for maximum goals and total programs. The SEA establishes minimum standards, the LEAs set maximum goals, and the local school fits these into programs that meet their students' unique needs. LEA testing programs should examine higher order skills. In establishing, implementing, and assuring minimum standards, the Department has a leadership responsibility to assure that maximum goals and unique needs are constantly and consistently emphasized and reinforced within our LEAs. The long term goal of striving for maximum goals has not at all disappeared in Louisiana, but a quality assurance program of mastery of minimum standards is a more immediate goal.

Louisiana has taken strong aggressive steps toward assuring higher quality education for its students. A thoughtful, well-planned, comprehensive program is being slowly and carefully developed. In this program basic skills testing and the establishment of mastery standards are key components. Great care has been taken to assure

FIGURE 6

DEVELOPMENT OF BASIC SKILLS TESTING PROGRAM

LOUISIANA LITERACY ASSESSMENT PROGRAM

STATE ASSESSMENT (ACT 621)

1979-80	1980-81	1981-82	1982-83	1983-84	1984-85
4-8-11	3-7-10	7-10	7-10	7-10	7-10
READING	READING	READING	READING	READING	READING
MATHEMATICS	WRITING	WRITING	WRITING	WRITING	WRITING
WRITING	MATHEMATICS	MATHEMATICS	MATHEMATICS	MATHEMATICS	MATHEMATICS

BASIC SKILLS TESTING (ACT 750)

LANGUAGE ARTS AND MATHEMATICS

DEVELOP	PILOT	IMPLEMENT	IMPLEMENT	IMPLEMENT	IMPLEMENT
2ND	2ND	2ND	2ND AND 3RD	2ND, 3RD, AND 4TH	2ND, 3RD, 4TH, AND 5TH
	DEVELOP	PILOT	PILOT	PILOT	PILOT
	3RD AND 4TH	3RD AND 4TH	5TH	6TH	7TH
		DEVELOP	DEVELOP	DEVELOP	DEVELOP
		5TH	6TH	7TH	8TH

that testing is not the only thing, but is instead a part of a comprehensive design.

The Involvement of Local Educators in the Development of State Basic Skills Tests[1]

Rebecca S. Christian
Louisiana Department of Education

Introduction

The Louisiana Basic Skills Testing Program, developed to meet the mandate of 1979 competency-based legislation will begin implementation during the 1981-82 school year. The testing program begins with the administration of a second grade criterion-referenced measure of minimum language arts and mathematics basic skills. Table 1 outlines the sequential grade implementation schedule of the Louisiana Basic Skills Program over a five-year period.

Minimum competency testing, as defined for the 1981 National Institute of Education (NIE) advisory evaluation hearing refers to programs mandated by a state or local body which have the following characteristics: "1) All students of designated grades are required to take paper-and-pencil tests designed to measure basic academic skills, life or survival skills, or functional literacy; 2) a passing score or standard for acceptable levels of student performance has been established; and 3) test results may be used to certify students for grade promotion, graduation, or diploma award; to classify students for or place students in remedial or other special services; to allocate compensatory funds to districts; to evaluate or to certify schools or school districts; or to evaluate teachers" (Thurston and House, 1981, p.

---

[1] A paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education. New York, March, 1982.

87). The Louisiana competency testing program meets almost all of the criteria defined and is anticipated to have a significant impact on students, educators, parents, special interest education groups, the State Legislature, the State Board of Elementary and Secondary Education and the concerned public. The student's test scores are to be the principal criteria in the promotion decision. Additionally, the student who does not achieve the established performance standard on either or both sections of the test (language arts or mathematics) must be provided compensatory/remedial education regardless of the promotion decision (Louisiana Revised Statutes 17:396, 397).

Other presentations in this symposium deal with the background and legislative history of the state testing program, issues of testing exceptional students, the impact of competency testing on urban school districts and the development and use of an item bank and item response theory in the development of the tests. The purpose of this paper is to discuss the Louisiana Department of Education's attempt to utilize the expertise of educator groups and to involve teachers, principals and curriculum supervisors in the development of a testing program that will have a major effect on their future job roles.

Test Specifications

Beginning in 1977, in accordance with Louisiana Revised Statutes 17:391.1-391.9, minimum standards were identified in the basic skills of reading, writing, and mathematics; and subsequently disseminated statewide. For each subject area an advisory council was established to develop the

standards. The advisory councils were composed of teachers, administrators, college professors, and specialists in that particular area. Each local school system had representation on at least one of the three advisory councils.

These standards were identified for teachers' instructional purposes, and list skills that students are expected to know at specified grade levels. In addition, minimum student performance standards enable teachers to focus their instructional programs upon those skills that are essential and clearly identify for students and teachers what is expected of them at certain grade levels.

Identification of Grade 2 Testable Skills In Language Arts and Mathematics

The grade level minimum standards, defined by the councils in 1977, provided the framework for the implementation of the Louisiana State Assessment Program. In 1979 these same basic skill competencies became the cornerstone for the development of the Louisiana Basic Skills Testing Program.

Basic skills indicated for mastery at the end of Grade 2 as outlined in the minimum standard documents, Minimum Standards--Maximum Goals for Reading, Grades 1-12 (Bulletin 1488), Minimum Standards for Mathematics, Grades 1-12 (Bulletin 1497), and Minimum Standards for Writing, Grades K-12 (Bulletin 1502) were identified as testable objectives for the Grade 2 State Basic Skills Testing Program.

Members of the previously appointed reading, writing, and mathematics

assessment advisory committees were asked to prioritize skills in the order of relative importance to the grade two curriculum. This compilation was used in a final determination of objectives to be measured on the Grade 2 Louisiana Basic Skills Test.

Based on the skills identified in the minimum standard documents and advisory recommendations, general test specifications and item specifications were developed by the Department of Education's staff.

The test specifications for the Grade 2 Basic Skills Test broadly describe physical and statistical characteristics, as well as general test content. Rules for determining the difficulty level of passages, acceptable readability levels and suggested sentence length were incorporated. The Harris-Jacobson Word List was adopted after a review of approved basal texts to determine the acceptability of multiple syllabled words.

Item specifications clarified the test content by specifying the manner in which grade level mastery skills, restated as testable objectives, would be measured.

A sample item specification appears in Appendix 1. Staff members of Louisiana State University and the Educational Testing Services, Princeton, New Jersey provided consultative services both prior to and during the development of the specifications. Two members of the ETS staff offered a final critique of the language arts and mathematics specifications.

Item Development and Validation

The Department of Education determined, prior to the beginning of the test development phase of the program, that outstanding Louisiana classroom

teachers possessed the skills to produce quality items for the Basic Skills Testing Program.

Classroom teachers were recommended by the superintendents of the local school systems to participate in an item writing workshop. The participants were selected from sixteen school systems, but they represented a cross section of lower elementary teachers statewide. The ETS consultants conducted an intensive training session on item writing procedures and rules. The Department of Education's staff coordinated the week-long, item development workshop.

Based on the general test specifications and item specifications, twenty items per objective were developed in this session. A review by the ETS staff consultants followed. The form used by item writers and reviewers appears as Appendix 2. Suggestions and revisions were incorporated as items were finalized for field-test purposes.

The Department of Education conducted a content validation review prior to finalizing field-test forms for the scheduled pilot in the spring of 1981. Key advisory council members from the elementary sections reviewed items for content validity. Additionally, they recommended, from a content perspective, ten items per objective to be included in the 1981 Louisiana Basic Skills Field Test.

ETS consultants conducted a final review of the test forms and assisted the Department of Education in the development of manuals to accompany each test form.

The Grade 2 Test Pilot

The Grade 2 Pilot was conducted in the spring of 1981. The Research Triangle Institute of Raleigh-Durham, North Carolina was contracted to design a scientific probability sample to be utilized in the Grade 2 Basic Skills Field Test. Every school system was involved to some extent in the field test. The sample involved approximately 12,000 students, a 20 percent sample of Louisiana public school second graders, and included 480 classrooms in 462 schools.

Ten items per objective were field tested in language arts and mathematics. Multiple forms, six language arts forms and six mathematics forms, were administered. For each subject area, ten core items appeared in each form to provide a sounder statistical basis for establishing reliability.

A teacher perception section appeared on every Grade 2 Basic Skills Field Test answer document. This section was to be filled out by the test administrator/teacher. The purpose of this section was to provide information to assist the Louisiana Department of Education in establishing minimum acceptable levels of performance for the 1982 Grade 2 Basic Skills Test. Each teacher who administered a mathematics pilot test was required to respond to the following item:

Consider the performance of this student in your class. How would you expect this student to perform on a fundamental Grade 2 mathematics test?

0 I would expect this student to pass.

0 I would not expect this student to pass.

0 I would be unable to predict this student's performance as clearly passing or failing on such a test.

Teachers who administered language arts tests responded to the same item, except the words "language arts" were substituted for the word "mathematics."

This information provided a research base, for contrasting group and borderline group analyses, and thereby allowed the Louisiana Department of Education to consider the effects of setting various cut-off scores in terms of the second grade population.

Item response data from the grade two field-test administration was used by the LDE to determine item difficulty and item discrimination. In addition, item response analysis was conducted in order to maintain equivalence of test forms.

The Basic Skills Planning Commission

In order to provide for the involvement of all local education agencies on a continuing basis in the development of the Basic Skills Testing Program, Grades 2-12, a statewide planning commission was appointed. The superintendents of local school systems designated two persons from their staffs to serve as members of the Basic Skills Planning Commission. In

addition, persons representing the special interests of exceptional students, limited-English proficient and non-English proficient students and minority-disadvantaged students were appointed to membership on the Commission. Duties of the Commission were defined as both immediate and long-range.

The primary task was to provide an intensive review of the content of the Grade 2 Basic Skills Test prior to finalization of the test instrument and the determination of the performance standard.

The Commission participated in a bias review to determine that test items were free of offensive sexual, cultural, social, and/or ethnic content or stereotyping. A second content validity review was conducted to assure the items were content valid. Subject area specialists on the commission determined if the content measured by the test was appropriate to the instruction students are actually receiving in Louisiana public schools.

Louisiana Revised Statutes 17:24.4E required the determination of a performance standard. Clearly the various approaches under consideration for standard setting called for the extensive involvement of classroom teachers, principals and curriculum supervisors.

Twenty-six members of the Basic Skills Planning Commission participated in an Angoff standard setting exercise. Thirty-one of the Angoff participants were grade two classroom teachers who had administered the pilot test.

The Angoff method is a judgmental standard-setting method. Judges are asked to conceptualize the "minimally accepted person." The judges then assign a probability to each item refelecting their estimate of the prob-

ability that the minimally acceptable person should correctly answer the item. The results are then totaled and divided by the number of judges to determine the average.

The participants were given an explanation of the Angoff standard setting method and were then asked to rate all items in a particular subject area. Judges had both the test items and the portion of the test administrator's manuals which referred to those items. These data were used in conjunction with the contrasting groups and borderline group data as input in the standard-setting process.

The second phase of the Basic Skills Planning Commission's task provided continued involvement by the local education acency in developing tests for other grade levels. During the spring of the 1980-81 school year the Department of Education initiated test development activities for grades three and four.

The Basic Skills Planning Commission was involved in the following tasks:

(1) Prioritization of the grade three and grade four domains and skills described in the minimum standard documents Minimum Standards--Maximum Goals for Reading, Grades 1-12 (Bulletin 1488), Minimum Standards for Mathematics, Grades 1-12 (Bulletin 1497), and Minimum Standards for Writing, Grades K-12 (Bulletin 1502). (Appendix 3 contains the review forms used in this this task.)

(2) Review of specifications developed by the Department of Education's

staff. (A sample of the review form appears in Appendix 4.)

(3) Designation of two persons from their local school system with expertise in language arts or mathematics to be trained in item writing techniques and to participate on item writing teams.

The item writing teams were representative of the sixty-six local school systems. Items produced were reviewed and recommended for inclusion in the 1981-82 Grade 3 and Grade 4 Field Test by the Basic Skills Planning Commission and the Educational Testing Service, Princeton staff consultant.

After the 1981-82 field testing, the Basic Skills Planning Commission will again participate in content and bias reviews for the finalization of tests at these grade levels. The Commission will provide the Department of Education with extensive local involvment in all phases of test development according to the implementation schedule in Table 1. The Louisiana Department of Education disseminated to local education agencies in the fall of the 1981-82 school year copies of the test and item specifications for the Grade 2 Basic Skills Test. Copies were allocated to the districts on a per school basis so that classroom teachers would have an opportunity to become familiar with the parameters of the test. Additionally, coordinated staff development activities at the state, regional, and local level have been conducted to help teachers and administrators focus on strategies for promoting students' mastery of the target skills through the use of state-approved curriculum guides.

The Louisiana Department of Education's staff will conduct a series of utilization workshops upon completion of the testing. The purpose of the workshops is to assist the central office staff in training teachers, principals, and supervisors to properly use the various reports the systems will receive as a result of the program. School systems will also receive direction in developing compensatory/remedial programs at the local level.

Summary

W. James Popham (1981) of UCLA, the pro-team leader of the NIE advisory evaluation hearing on minimum competency testing, defined seven characteristics of a high quality Minimum Competency Testing Program as follows: "1) Conscientious selection of competencies; 2) Appropriate competency tests; 3) Teaching/testing congruence; 4) Multiple testing opportunities; 5) Adequate phase-in time; 6) Sensible setting of standards; and 7) Coordinated staff development" (p. 90).

The foundation for the Louisiana Basic Skills Program was provided as classroom teachers, principals, curriculum supervisors and higher education personnel defined minimum standards in the basic skill subjects. Educator groups continued to be actively involved in all phases of test development.

This involvement continues to include participation in objective, specification and item reviews, item writing sessions and participation in standard-setting activities. Minimum standard publications, curriculum guides and item specifications have been made available to teachers in the field. Staff development and inservice activities have been conducted to assist teachers in utilizing these tools to improve instruction and teaching strate-

gies.

The Louisiana Department of Education has attempted through the extensive involvement of educators in the field to develop a quality minimum competency testing program that will yield beneficial outcomes for public school students.

TABLE 1

LOUISIANA LITERACY ASSESSMENT PROGRAM

STATE ASSESSMENT (ACT 621)

1979-80	1980-81	1981-82	1982-83	1983-84	1984-85
4-8-11	3-7-10	7-10	7-10	7-10	7-10
READING	READING	READING	READING	READING	READING
MATHEMATICS	WRITING	WRITING	WRITING	WRITING	WRITING
WRITING	MATHEMATICS	MATHEMATICS	MATHEMATICS	MATHEMATICS	MATHEMATICS

BASIC SKILLS TESTING (ACT 750)

LANGUAGE ARTS AND MATHEMATICS

DEVELOP	PILOT	IMPLEMENT	IMPLEMENT	IMPLEMENT	IMPLEMENT
2ND	2ND	2ND	2ND AND 3RD	2ND, 3RD, AND 4TH	2ND, 3RD, 4TH, AND 5TH
	DEVELOP	PILOT	PILOT	PILOT	PILOT
	3RD AND 4TH	3RD AND 4TH	5TH	6TH	7TH
		DEVELOP	DEVELOP	DEVELOP	DEVELOP
		5TH	6TH	7TH	8TH

M-2-8

GRADE: Second

SUBJECT: Mathematics

STANDARD: The student will demonstrate a knowledge of whole number operations by adding a two-digit number and a one- or two-digit number, no regrouping.

SKILL: Add a two-digit number and a one- or two-digit number, no regrouping.

CLARIFICATION OF SKILL: In order to demonstrate a knowledge of this skill, the student will add a two-digit and a one- or two-digit number and select the correct response from the options given.

MINIMUM STANDARD REFERENCE: Whole Number Operations, III-6, 8

Stimulus Attributes

A. Format:
Addition problem written:
1. vertically or
2. horizontally

B. Content:
A two-digit number and a one- or two-digit number

C. Directions: (To be read by the test administrator.)
Add the two numbers at the beginning of the row. Look across the row and find your answer. Fill in the space under your answer.

Response Attributes

A. Format:
Four numbers listed horizontally.

B. Options:
1. The correct response
2. The three distractors may include the number that is:
   (a) The sum of ones column only
   (b) The difference of the two numbers
   (c) The sum of the three/four digits
   (d) One more/less than the sum

Sample Items

Directions: (To be read by the test administrator.)
Add the two numbers at the beginning of the row. Now look across the row and find your answer. Fill in the space under your answer.

1.

26 +72	88 0	89 0	97 0	98 0

2.

54 + 2 = ☐	52 0	56 0	74 0	76 0

A Second Grade Language Arts Item Specification used in the development of the Louisiana Basic Skills Item Bank

Test: Second Grade Basic Skills Test

Subject: _____ Classification: _____

Submitted by: _____ Date: _____ Item type: MC

(Type or print item here)

_____ (Indicate key on back)

Reviewers' Comments:

KEY_____

Source (for quoted material, graphs, cartoons, etc.):

Appendix 3

Domain Prioritization Form

DOMAIN PRIORITIZATION FORM

-LANGUAGE ARTS-

This form is divided into two sections. Each section contains a list of language arts domains for a given grade for which testable objectives can be found in the minimum standard documents. In parentheses beside each domain is the number of testable objectives within the domain.

Read the list of Grade 3 language arts domains. In the blank to the right of each domain, indicate the relative importance of the domain on a fundamental Grade 3 language arts test by writing the percentage of items on such a test which should measure the domain. The sum of the percentages you assign to Grade 3 language arts domains must equal 100 percent.

Repeat the process for Grade 4 domains.

GRADE THREE

DOMAIN	RECOMMENDED WEIGHT (by %)
Vocabulary (2)	_____
Phonetic Analysis (10)	_____
Structural Analysis (7)	_____
Comprehension (7)	_____
Study Skills (6)	_____
Types and Forms of Writing	
Narration (1)	_____
Exposition (2)	_____
Writing Skills	
Handwriting (2)	_____
Spelling (5)	_____
Capitalization (6)	_____
Punctuation (5)	_____
Language Structure (5)	_____
Organization (1)	_____
TOTAL (must = 100%)	_____

GRADE FOUR

DOMAIN	RECOMMENDED WEIGHT (by %)
Vocabulary (12)	_____
Phonetic Analysis (6)	_____
Structural Analysis (7)	_____
Comprehension (12)	_____
Study Skills (9)	_____
Types and Forms of Writing	
Exposition (1)	_____
Writing Skills	
Spelling (4)	_____
Punctuation (2)	_____
TOTAL (must = 100%)	_____

DOMAIN PRIORITIZATION FORM

-MATHEMATICS-

This form is divided into two sections. Each section contains a list of mathematics domains for a given grade for which testable objectives can be found in the minimum standard documents. In parentheses beside each domain is the number of testable objectives within the domain.

Read the list of Grade 3 mathematics domains. In the blank to the right of each domain, indicate the relative importance of the domain on a fundamental grade three mathematics test by writing the percentage of items on such a test which should measure the domain. The sum of the percentages you assign to grade three mathematics domains must equal 100 percent.

Repeat the process for grade 4 domains.

GRADE THREE	
DOMAIN	RECOMMENDED WEIGHT (by %)
Numeration (8)	_____
Whole Number Operations (19)	_____
Fractions and Operations (2)	_____
Relations and Functions (5)	_____
Measurement and Estimation (8)	_____
Geometry (1)	_____
Problem-Solving (4)	_____
TOTAL (must = 100%)	_____

GRADE FOUR	
DOMAIN	RECOMMENDED WEIGHT (by %)
Numeration (5)	_____
Whole Number Operations (10)	_____
Fractions and Operations (2)	_____
Relations and Functions (6)	_____
Measurement and Estimation (3)	_____
Geometry (3)	_____
Problem-Solving (1)	_____
TOTAL (must = 100%)	_____

Objective Prioritization Form

GRADE 3 MATHEMATICS

DOMAIN: NUMERATION

Read the following objective statements. In the blank to the left of the statement rank order the statements to reflect their importance on a fundamental basic skills test of this domain. Mark the most important objective with a "1", the second most important objective with a "2", and so on until all objectives have been ranked.

1. Number Words

_____ a. Students will identify number words for multiples of ten.

_____ b. Students will identify number words eleven through nineteen.

_____2. Write Numbers

Students will write to one hundred by twos.

_____3. Read and Write Numbers

Students will read and write numbers through one thousand.

_____4. Place Value

Students will recognize place value, hundreds.

_____5. Ordinal Numbers

Students will use ordinal numbers through nineteenth.

6. Sequence

_____ a. Students will supply missing numbers in a sequence through one hundred.

_____ b. Students will write the number that comes before (after) a given number. (1-99)

In your professional opinion, should any of these objectives be combined for the purpose of a basic skills test? If so, which ones?

Appendix 4

Item Specification Review Form

Item Specification[1]

Review Form

Item Specification:_____ Reviewer:_____ Date:_____

Please read the item specification carefully. Next answer the eight questions below. Staple (or clip) your copy of the item specification to this review form after you have answered the questions.

Thank you for your comments.

SKILL

1. Does the "skill section" provide sufficient details to give a reader an indication of the behaviors defined by the item specifications? (Circle One)

   (a) Yes
   (b) Yes, with reservations
   Please explain:

   (c) No

2. How might the "skill section" be revised to improve its clarity?

Sample Directions and Test Item

3. Can the test directions be revised to improve their clarity? (Circle One)

   (a) Yes, (please write your comments on the back of this form)
   (b) No, they are clearly written

4. Do you feel the item format is the best to ensure that examinee answers will measure the behaviors defined by the general description? (Circle One)

   (a) Yes
   (b) Yes, with reservations
   Please explain:

   (c) No. Which item type would be best?

5. Will the sample test item provide a "good model" for item writers preparing items to measure the item specification? (For example, does the item include the desired number of answer choices, and is the vocabulary appropriate for the intended group of examinees?) (Circle One)

   (a) Yes
   (b) Yes, with reservations
   Please explain:

   (c) No

STIMULUS ATTRIBUTES

6. Do you have any suggestions for revising and/or extending the content defined by the item specification? Please write comments on the back of this form. (Your suggestions could include: (a) deletions of specific content, (b) additions to the content, (c) rewrites for clarification).

CHARACTERISTICS OF RESPONSE ATTRIBUTES

7. Do you have any suggestions for revising and/or extending the characteristics of possible answers? Please write comments on the back of this form.

---

[1]Adapted from Hambleton, R. K. and Eignor, D. R. A practitioner's guide to criterion-referenced test development, validation, and test score usage. Laboratory of Psychometric and Evaluative Research Report No. 70. Amherst, MA: School of Education, University of Massachusetts, 1979.

References

Thurston, P. & House, E. R., The NIE adversary hearing on minimum competency testing, Phi Delta Kappan , 1981, 63, 87-89.

Popham, W. J. The case for minimum competency testing, Phi Delta Kappan , 1981, 63, 89-91.

IMPLICATIONS OF THE LOUISIANA BASIC SKILLS TEST FOR SPECIAL EDUCATION STUDENTS

by

Ellen B. Gillespie, Ph.D.
Louisiana State Department of Education
Baton Rouge, Louisiana

Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education. New York, 1982.

IMPLICATIONS OF THE LOUISIANA BASIC SKILLS TEST FOR SPECIAL EDUCATION STUDENTS

The development of the minimum competency testing movement in the last ten years was originally a reaction to public outcry over test scores indicating declining skills of high school graduates. As minimum competency testing programs have been implemented, however, they have developed into more comprehensive programs, often using test data for promotional decisions for all students. The many state and local programs vary along several dimensions, including program goals, competencies assessed, grade levels included, and provisions for special populations.

The issue of inclusion of special education students in such testing programs has been affected by federal legislation. Public Law 94-142, the Education of All Handicapped Children Act, was enacted in 1975 and assures that handicapped children receive a free appropriate public education. An additional federal law, Section 504 of the Rehabilitation Act of 1973, prohibits recipients of any federal funds from discriminating against handicapped persons solely on the basis of their handicap. These two landmark pieces of legislation combine to assure the provision of educational and related services to all handicapped children, regardless of the severity of their handicapping condition.

The issue, then, of inclusion or exclusion of special education students from statewide testing programs is covered by these two laws. Contingent upon performance on tests, many state and local education agencies award diplomas, promote students from grade to grade, and allocate resources for compensatory or remedial programs. Special education students may not be categorically excluded from receiving funding or privileges available to other students as a result of classification as a student with special needs.

From a more global perspective, special education students may not be excluded from possible long-term improvements in curriculum and instruction which may accrue from the implementation of a competency-based education program.

The legislation mandating Louisiana's Basic Skills Test, therefore, was consistent with federal legislation in specifically including special education students. The development of policy alternatives for the testing of these students, however, was a complex process. The primary purposes of this paper will be: (a) to review major issues involved in implementing a statewide testing program which includes special education students and (b) to describe Louisiana's response to these issues.

Five issues will be discussed: (1) inclusion/exclusion of special education students, (2) development of alternative standards for students not tested, (3) test procedural and format modifications, (4) issues involved in research based on the test performance of special education students, and (5) involvement of consumer advocacy organizations.

Inclusion/Exclusion of Special Education Students

As previously stated, current federal legislation (as well as state legislation and litigation) has established the participation of special education students in educational programs. Arbitrary exclusion of special students from testing programs would be inconsistent with these laws. Some minimum competency testing programs, however, have chosen to allow exclusion of students in certain special education categories, particularly students whose intellectual abilities are severely impaired. Other programs have allowed for "exemptions" for some students on an individual basis, with the decision for exemption usually requiring parental permission and having an effect on receipt of a diploma or on promotion.

In developing policies for including special education students, the Louisiana Department of Education avoided exclusion or exemption based on special education classification. Instead, the decision as to whether a child would be tested would be made by the child's individual education program (IEP) committee. P.L. 94-142 provides for the annual development of an IEP for each child by school personnel, parents, and the child, whenever appropriate. Each child's educational program is to be considered in the IEP meeting to determine if the child is addressing state minimum skills, and, therefore, should be tested on these skills. Students determined by the committee to be addressing state minimum skills in their educational program (although not necessarily to have mastered such skills,) are said to be in "specially designed regular instruction." These students are tested as a part of the Basic Skills Testing Program. Students who are not addressing state minimum skills are said to be in an "alternative to regular placement." These students are not included in the Basic Skills Test Program. A standard IEP form is developed by the State Department of Education and used for all IEPs in the state. Space to designate specially designed regular instruction or alternate to regular placement is included on the form. Decisions as to participation in the testing program are made annually. As an additional safeguard, parents whose children are in an alternative to regular placement must sign a statement on the IEP form indicating that they understand that this placement may mean the child will not receive a regular diploma. Failure to obtain the signature means that the child will participate in the testing program.

A distinct advantage of this policy over the exclusion of some categories of special education students is the individualization of the process. Input from IEP committee members who know the child and his school program make

the decision annually. One disadvantage of this policy alternative is inconsistency in definition of the two terms, specially designed regular instruction and alternative to regular placement. Some misunderstanding exists among practitioners concerning a distinction between resource room (or "mainstreamed") students and those in a self-contained class. Placement in a particular type of class actually has no bearing on the decision of including or excluding a child from the testing program. A second possible problem area arises concerning the issue of informed consent. Research concerning IEP meetings (Gilliam, 1979; Goldstein, Strickland, Turnbull, and Curry, 1980) has indicated that parents are often passive participants in such meetings. It is difficult to determine whether parents understand the long-term implications of making such a decision.

Additional issues arise in developing policy for testing special education students in ungraded classes. This is a common situation with most self-contained classes. Students are not actually assigned to the second grade or third grade, but may be in a class for elementary mildly mentally retarded children.

In developing strategies for dealing with the problem of testing level for special education students in ungraded classes, the Louisiana Department of Education studied the following possibilities: (1) excluding students in ungraded classes; (2) including only those students whose chronological age would indicate they would ordinarily be placed in the second grade; (3) allowing off-level testing and (4) requiring teachers to determine the grade level at which the student was currently functioning. The various strategies were considered in a series of meetings with state Accountability and Special Education staff, as well as meetings with local special education supervisors and test coordinators. The groups consulted were concerned that criteria

for testing students in ungraded classes not disallow any student who might benefit from participation in the Basic Skills Test. At the same time, it was not seen as beneficial for a student addressing but not having mastered second grade skills to continue to take the second grade test year after year.

The compromise position reached by Department staff resulted in the development of the following criteria for testing special education students in ungraded classes: (1) placement in specially designed regular instruction according to the IEP, (2) teacher judgment as to whether the student's IEP indicates that he is addressing second grade state minimum skills, and (3) a chronological age range of 4 years. For the Grade 2 test for example, a student in an ungraded class must be 7, 8, 9 or 10 years of age.

The decision as to the grade level reflected by the student's IEP is made at the school level. While the use of a range of chronological ages allows for developmental delay and difficulties in learning, this system does not require that schools determine how many 14-year-olds, for example, are addressing second grade skills. It is very likely that students older than 10 and still addressing second grade skills will be placed in an alternative to regular placement in the IEP meeting.

In summary, the issue of including or excluding special education students is not a simple one. The diverse nature of special education dictates that complex solutions are often necessary. The Louisiana Department of Education has chosen to develop policies which allow for maximum individualization in determining whether a student should be tested.

Development of Alternative Standards

The existence of a dichotomy concerning Louisiana students' participation in the Basic Skills Test indicates the need for a means of evaluating

those students not in specially designed regular instruction. Often these students are addressing non-academic areas as major emphases in their school program. Motor skills, self-help training, and basic communication skills may comprise a student's program.

In order to establish criteria for promotion and graduation for this group of students, the Office of Special Education of the Louisiana Department of Education organized task forces of skilled teachers to develop alternative standards for special education students. Two sets of skill sequences have been developed and are being pilot tested this year: Alternative Standards for Mildly Handicapped Students and Alternative Standards for Moderately, Severely, and Profoundly Handicapped Students. The documents include skill sequences for elementary and secondary students in all curriculum areas.

Use of the alternative standards is then the basis for the awarding of the state Certificate of Achievement, an alternative to the regular diploma. In order to receive the Certificate of Achievement at the completion of his school career, a student must have met attendance requirements and must have completed a minimum of 70% of his IEP objectives. IEP objectives are to be based on the skill sequences from the appropriate alternative standards document. Teacher judgment is used to determine completion rate for objectives.

Strengths of this approach to alternative standards include: (1) the extension of the accountability program to all students, not just those studying reading and mathematics, and (2) development of a mechanism for a state-approved certificate for special education students not addressing state minimum skills. The certificate is not awarded simply on the basis of years enrolled in school, but implies mastery of an individually designed program of education.

Test Procedural and Format Modifications

The development of test modifications is another issue of major concern in implementing a statewide testing program including special education students. While test modifications should not be so extensive as to give an unfair advantage to special education students, it is apparent that a student with a visual impairment will need a Braille or large-print test, that the student with poor fine motor skills will need assistance in completing a machine-scorable answer sheet, and that a student with a limited attention span may need extra time or multiple sittings for test completion. The key to effective test modifications is accommodation for a child's handicap without invalidating the test as a measure of basic skills.

The difficulty in developing options for test procedural or format modifications is the lack of research in this area. McKinney and Haskins (1980) studied the extent of utilization of modifications on the North Carolina Minimum Competency test. One difficulty with this study was the analysis of use or non-use of modifications, rather than study of the effects of specific modifications on test performance. An extensive assessment of test modifications for learning disabled students has been completed by the Florida Department of Education (Beattie and Algozzine, 1981). Further research is needed for other exceptionalities, however, including elementary-level students likely to be involved in minimum competency programs used for promotional purposes.

Currently, Louisiana allows most of the test modifications included in other minimum competency testing programs. Test modifications for the 1981-82 Basic Skills Test include the following:

1. Braille edition.
2. Large-print edition

3. Answers recorded by test administrator
4. Extended time allowed for test administration
5. Instructions given in sign language
6. Student allowed to mark responses in the test booklets
7. Test administered individually or in small groups
8. Directions repeated
9. Math test read to students

Louisiana's policy related to modifications is unique in that test modifications are not limited to students with particular exceptionalities, but may be used by any special education student as needed. A learning disabled or mentally retarded student who might be easily distracted by a page containing many test items, lines, or pictures could use a large-print test booklet with a simplified page layout. Test modifications are selected in the IEP meeting, thus allowing for input from parents, therapists, and teachers.

One possible concern related to test modifications is rate of utilization. Some modifications, such as individual or small group administration, may pose logistical problems for practitioners. One possibility is that special education students tested by regular teachers in a mainstreamed homeroom will be less likely to be allowed modifications. A second possibility is that timing of fall state-level data collection for large-print tests makes it difficult to estimate the number of students who might need such a test format. Research concerning utilization of modifications and effects of each of the modifications on test performance will be conducted each year by the Louisiana State Department of Education.

Basic Skills/Special Education Research

Despite the critical need for research into the effects of the Basic Skills

Test on special education students, experience in field testing has led Department of Education personnel to note several difficulties in conducting such research:

1. Low incidence handicaps

Field testing of the Grade 2 Basic Skills Test involved a 20% statewide sample. The sample was selected by using classroom teachers' names. Of the approximately 11,000 students participating in the field test, 490 were special education students. An additional 19 students in residential schools for sensorially impaired persons participated in a special pilot study. At Grade 2 many special education students have not yet been identified. Even considering the lower rate of identification, however, special education students were underrepresented in the field test sample. The result of the low proportion of participation by special education students in the pilot was an inadequate number of students in most special education classifications for purposes of analysis. In order to alleviate the problem this year a special pilot of students in higher incidence categories (learning disabled, mildly mentally retarded, and slow learners) is planned. The problem of numbers will remain, however, with exceptionalities such as visual impairments, hearing impairments, and orthopedic handicaps. In field testing, when multiple test forms are used, the problem becomes even more complex.

2. Possibility of misclassification

A second difficulty in conducting testing research with special education students is the possibility of misclassification. McKinney and Haskins (1980) noted that misclassification was a significant problem in the North Carolina Competency Testing Program. Two kinds of misclassifications can occur: coding errors or actual misclassification as to exceptionality. In Louisiana, test administrators are required to code the student's special education

classification on the test answer sheet. Possible sources of error include lack of familiarity with the student's IEP and confusion over which of several handicapping conditions may be the primary handicap. Test administrators are instructed that classification must coincide with information on the student's IEP. Monitoring of a sample of students to determine misclassification rate is planned this year.

A more serious problem occurs when actual misclassification has taken place. Although federal and Louisiana state legislation has numerous safeguards concerning pupil appraisal and placement, some error does occur and must be considered in conducting research.

3. Statistical analysis

The appropriateness of statistical procedures commonly used for reporting test scores of regular students is another area of concern. The average percent correct (APC) is the statistic which has typically been used for score reporting of test results in Louisiana. An advantage, then, of reporting performance of special education students using the APC is familiarity with this type of score. Reporting for regular education students on the Louisiana State Assessment is based on the APC at the state, parish, school, and individual student level. A disadvantage of this form of reporting, however, is the amount of information omitted by the use of the APC alone. The reporting of an average score or APC does not give information about the distribution of scores. Within a special education classification, such as learning disabled, for example, the dispersion among scores is not indicated by the mean. A second disadvantage in relying on APC reporting alone for special education students is that the mean is strongly affected by extreme scores when a distribution is very skewed and numbers are small. Distributions of scores within special education

categories are often skewed distributions. A possible solution for this problem is the reporting of a median and semi-interquartile range in addition to a mean for special education categories.

4. Political implications of score reporting

One additional difficulty in dealing with research related to the testing of special education students concerns score reporting. The possibility of aggregating scores of special education students and regular students is one alternative, especially at the state and district levels. In situations where comparisons are made by the news media among school districts this alternative is not considered desirable by school administrators. School districts including more special education students in testing would be likely to have lower district-level mean scores.

Separate score reporting then becomes necessary for special education students. In Louisiana there are 15 special education classification codes used for statewide testing. Because a report which furnishes information about all special education students combines such diverse groups as gifted/talented, visually impaired, and autistic students, separate reports are needed for each exceptionality at each level. The production and distribution of so many reports then becomes a logistical problem.

One interesting point concerning score reporting for special education students is that there are certain exceptionalities which may not be treated separately from regular education students because of political implications. In Louisiana, a compromise was reached by reporting scores of gifted/talented, hospital/homebound, and speech impaired students with the scores of regular students. School administrators do not wish to have these groups treated separately.

The politics of score reporting is a major consideration in any minimum

competency testing program. The additional considerations of special education reporting make this area even more complex.

Involvement of Consumer Advocacy Organizations

Since the 1960s the growth of consumer organizations representing the interests of handicapped persons has been phenomenal. Parent groups are skilled as lobbyists, service providers, and as public information specialists.

The relationship between consumer groups and education personnel involved in test development and implementation is an area which has typically been given little attention. Such groups as the Association for Children with Learning Disabilities, the Association for Retarded Citizens, and the Coalition for Citizens with Disabilities, however, are knowledgeable about educational programming and have concerns about the implications of state testing programs for the students they represent.

In Louisiana, the Department of Education has made involvement of consumer groups a priority with the implementation of the Basic Skills Test. Activities which have been accomplished with these groups include:

- Distribution of printed information concerning the Basic Skills Test (test modifications, definition of terms, criteria for testing students in ungraded classes, parent activities for teaching basic skills)
- Involvement of group representatives in Bias Review Workshops
- Presentations by State Department of Education Staff at numerous state and local advocacy organization meetings to explain testing policies and procedures

The Louisiana Department of Education has worked to establish open communication with consumer advocacy organizations. While all members of the various groups are not advocates for the Basic Skills Test, willingness to share information and accept suggestions has fostered an improved relationship between the State Department of Education and consumer groups concerning the test.

Summary

In summary, the Louisiana Department of Education is involved in the implementation of a Basic Skills Test which will serve as a measure of skills for regular and special education students. Policy development has focused on individualizing decisions concerning test participation and modifications. Alternative standards have been developed for special education students not addressing state minimum skills. State Department of Education staff are working to solve difficulties in conducting research and reporting test scores of special education students. Consumer organizations are being involved in test development and implementation.

The application of minimum competency testing programs to special education students poses a variety of difficult problems. The implications of each policy and procedure must be considered in light of students with a range of physical and intellectual abilities. The outcome of careful consideration of these issues in planning testing programs will be the development of policies which assure a test which is fair for all students involved.

References

Beattie, S. and Algozzine, B. Assessment of minimum competency in grades three and five: An analysis of modifications to Florida's State Student Assessment Test-I (Final report draft - Department of Education Contract #080-187). Gainesville, FL: University of Florida, April 1981.

Gilliam, J. E. Contributions and status rankings of educational planning committee participants. Exceptional Children, 1979, 45, 466-467.

Goldstein, S., Strickland, B., Turnbull, A. P., and Curry, L. An observational analysis of the IEP conference. Exceptional Children, 1980, 46, 278-286.

McKinney, J. D. and Haskins, K. G. Performance of exceptional students on the North Carolina Minimum Competency Test. Chapel Hill, NC: University of North Carolina at Chapel Hill, 1980.

The Development and Use of An Item Bank for a State Testing Program[1]

Craig N. Mills
Louisiana Department of Education

Introduction

During the 1981-82 school year, a Basic Skills Testing Program is being implemented in Louisiana. The program is beginning with a second grade test this year. Each year an additional grade is to be incorporated into the program. Any student who does not achieve the established performance standard on either section of the test (language arts or mathematics) will receive remediation. Additionally, the results of the tests are to be the principal criterion which influences promotion. Clearly, the Basic Skills Tests are going to have a significant impact on education and educators in Louisiana.

Several issues arise when a testing program has important consequences such as those of the Louisiana Basic Skills Tests. Other presentations in this symposium deal with some of those issues (testing special education students, involvement of local districts and personnel in test development, and the effects of state promotional testing on local districts). The purpose of this paper is to provide a discussion of the development of an item bank and the application of item response theory in the development and use of the bank.

[1] A paper presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education. New York, 1982.

During the development of the second grade test, it became clear that there would be a need for annual revision of the Basic Skills Tests. One purpose for annual revision is to maintain test security. Therefore, the decision was made to write and pilot many more items than were needed in the final test. This decision was made in order to (1) identify high quality items for possible use on the test, (2) identify items in need of revision and re-piloting and (3) create an item bank containing a pool of items of sufficient quality to be used for annual test revision. Items were piloted in the spring of 1981. A total of 162 language arts and 260 mathematics items were included in 12 pilot test forms. Following administration of the pilot test, the items were placed in an item bank. The purpose of this paper is to provide a discussion of item banking as a tool in the development and maintenance of a state basic skills testing program.

The remainder of this paper is divided into three sections. In the first section a brief introduction to item banking is presented. In the second section the process by which items were calibrated into the bank is explained. In the third section some applications of the item bank are briefly described.

Item Banking

Basically, item banking refers to a process by which items are collected, monitored, and stored in a computer with relevant information (such as content codes and statistical information) in such a manner as to facilitate retrieval of the items for test construction (Shoemaker, 1976; Thorndike, 1971). Although it is possible to

develop and manage item banks without the aid of a computer, usually computers are used for at least some of the functions of the bank. Item banks also include mechanisms for classifying items, introducing new items, editing and updating existing items, and deleting items which are no longer useful (Prosser, 1974).

The initial steps in the development of an item bank are similar to the initial steps in the development of a test. Tinkelman (1971) has provided an excellent discussion of the steps involved in the development of an objective test. A more detailed discussion, specific to criterion-referenced test development, was provided by Hambleton and Eignor (1979). An eleven-step process for the development of a criterion-referenced item bank is presented in Figure 1. Each step will be briefly discussed below.

1. Development of Item Bank Specifications

The first task in the development of an item bank is to define the purpose and scope of the bank. Item banks can be designed for purposes as varied as the production of many different tests covering a wide breadth of content to meet the divergent needs of many clients, to maintenance of an item pool in a specific content area for use by only the developer of the bank in production, and revision of a single test. It is important to accurately specify the purpose of the bank at the initiation stage of the project. Improper specification may lead to a design that is too restrictive or one that is grandiose and requires unnecessary effort to sustain it. Among the issues to be discussed in the development of the item bank specification are:

1. Development of Item Bank Specifications
2. Preparation of Item Specifications
3. Writing Items Matched to Specifications
4. Preliminary Assignment of Items to the Item Bank
5. Preliminary Review of Items
6. Pilot Trial of Items
7. Additional Editing of Items
8. Additional Pilot Trials (as Needed)
9. Selection of Items From the Bank for Test Development
10. Update of Item History and Performance
11. Periodic Review and Modification of Item Bank

Figure 1. Steps in the Development and Maintenance of an Item Bank

1. General purposes of the item bank
2. Breadth of content to be included in the bank
3. Selection or preparation of objective statements for which test items are to be written
4. Item type or types to be included in the bank
5. Determination of types of information to be collected and stored in the bank (content codes, item statistics, item source, history of use, selection criteria, etc.).

(2) Preparation of Item Specifications

The preparation of item specifications (also called amplified objectives or domain specifications) is an important step. The importance of these specifications has been noted elsewhere (Popham, 1978; Hambleton and Eignor, 1979). Basically, an item specification is an expansion and clarification of an objective statement. It serves as an explicit guide for item writers as they develop test items to be placed in the bank. If the item bank serves many clients, these item specifications can be of great value to them in the selection of skills to be tested with items from the bank. The item specifications may be provided to a client who can then select the skills that match curricular emphasis. Actual item selection may, however, be left to a computer.

Examples of two item specifications from the Louisiana Basic Skills Testing Program are shown in Figures 2 and 3. Each specification includes (1) an objective statement, reference, and clarification, (2) an explanation of acceptable item content, (3) rules for writing

M-2-8

GRADE: Second

SUBJECT: Mathematics

STANDARD: The student will demonstrate a knowledge of whole number operations by adding a two-digit number and a one- or two-digit number, no regrouping.

SKILL: Add a two-digit number and a one- or two-digit number, no regrouping.

CLARIFICATION OF SKILL: In order to demonstrate a knowledge of this skill, the student will add a two-digit and a one- or two-digit number and select the correct response from the options given.

MINIMUM STANDARD REFERENCE: Whole Number Operations, III-6, 8

Stimulus Attributes

A. Format:
Addition problem written:
1. vertically or
2. horizontally

B. Content:
A two-digit number and a one- or two-digit number

C. Directions: (To be read by the test administrator.) Add the two numbers at the beginning of the row. Look across the row and find your answer. Fill in the space under your answer.

Response Attributes

A. Format:
Four numbers listed horizontally.

B. Options:
1. The correct response
2. The three distractors may include the number that is:
   (a) The sum of ones column only
   (b) The difference of the two numbers
   (c) The sum of the three/four digits
   (d) One more/less than the sum

Sample Items

Directions: (To be read by the test administrator.) Add the two numbers at the beginning of the row. Now look across the row and find your answer. Fill in the space under your answer.

1.

26 +72	88 0	89 0	97 0	98 0

2.

54 + 2 = ☐	52 0	56 0	74 0	76 0

Figure 2: A Second Grade Language Arts Item Specification used in the development of the Louisiana Basic Skills Item Bank

LA-2-11

GRADE: Second Grade

SUBJECT: Language Arts

STANDARD: The student will demonstrate knowledge of study skills by alphabetizing to the first letter.

SKILL: Alphabetizes to the first letter

CLARIFICATION OF SKILL: In order to demonstrate an understanding of this skill, the student will select a word that comes first in alphabetical order.

MINIMUM STANDARD REFERENCE: Reading, Study Skills, Grade 2-A-1

Stimulus Attributes

A. Format: No stimulus is given. Only responses are listed.

B. Directions: Look across the row at each of the four words. Find the word that comes first in alphabetical order. Fill in the space under the word that comes first in alphabetical order. (To be read by the Test Administrator.)

Response Attributes

A. Format: Four words taken from the Harris-Jacobson Core Word List (second level or lower) arranged horizontally

B. Options:

1. The correct response is the word that comes first in alphabetical order.

2. The three distractors should begin with different letters of the alphabet.

Sample Item

Directions: Look across the row at each of the four words. Find the word that comes first in alphabetical order. Fill in the space under the word that comes first in alphabetical order. (To be read by the Test Administrator.)

cat	fish	apple	dog
O	O	O	O

Figure 3: A Second Grade Language Arts Item Specification used in the development of the Louisiana Basic Skills Item Bank

63

distractors, and (4) a sample test item. Item specifications should be reviewed by content experts to assure that they are sound.

(3) Writing Test Items Matched to Item Specifications

Although item specifications can be a substantial aid to item writers, item writing is still a difficult task. Item writers must exercise creativity without exceeding the constraints of the item specifications. If a single large item writing effort is to occur, rather than periodic small item writing projects, it is especially important that item writers be well-trained and understand the task before them.

(4) Preliminary Assignment of Items to the Item Bank

Once items have been written and prepared for review, they can be assigned identification and content codes and placed in the bank. These items are not, however, operational and should be so labelled. An "item clearance" variable may be recorded which has levels indicating the current status of the item. For example, an "R" might indicate that the item is at the review stage, a "P" that the item has been reviewed and is ready for piloting, an "E" might mean that the item is to be edited and repiloted and an "O" would signify that the item is operational and ready for use. As additional stages in item development are completed, relevant information (item statistics, history of use, etc.) can be integrated into the bank.

(5) Preliminary Review of Test Items

Two types of review are helpful in the development of test items. First, if item writers are working in teams, they should review one

another's work periodically. This allows them to benefit from one another's insights, identify good and poor algorithms for generating distractors, and identify areas where they may be violating the guidelines of the item specifications.

A second review, conducted by persons not involved in item writing, is also recommended. Content experts who review test items should be familiar with item construction techniques. Based upon these reviews items or even item specifications may be revised, discarded, or totally rewritten.

(6) Pilot Trial of Test Items

Content and technical reviews serve an important function in assessing item quality. Even the best reviews will, however, fail to detect many subtle flaws in items. These flaws will often become apparent when item analysis is conducted with pilot data. Item analysis results can point to items which (1) have item difficulties substantially different from those of other items that measure the same objective, (2) have negative or low item discrimination values, or (3) for which one or more distractors are attracting more high ability students than low ability ones. All of these indicate a need for further item review. While the item may still be cleared for use after review or revision and repiloting, these statistical reviews are valuable for identifying suspect items. Item analysis is also useful for identifying distractors which are attracting few examinees and which might be replaced with other distractors. This review can also be helpful in identifying possible new items or distractors.

Pilot tests can also serve a function in the integration of items into the bank. As will be explained later, item response theory is an useful tool in the development and use of an item bank. Pilot test results can be used for item calibration for both initial item bank development and, if the pilot is properly designed, for calibration of new items to be added to an existing bank.

(7) Additional Editing of Test Items

Based on the pilot test analysis, as described above, changes may be made to items or (occasionally) to item specifications following pilot testing.

(8) Additional Pilot Trials as Needed

If changes are made to an item it is, in effect, a new item. As such, the statistics which were gathered during the initial pilot no longer apply. The extent to which an item must be altered to be considered a new item in need of a pilot trial is not clear. The safest policy is one in which any change, no matter how minor, is grounds for additional pilot testing.

(9) Selection of Test Items from the Item Bank for Test Development

Once the bank has been stocked with a collection of items cleared as operational, it can begin to serve its clients. A variety of selection designs are possible, depending upon the purpose of the bank. These range from choosing items from a catalog to sophisticated algorithms for computer controlled item selection (Stodola, 1974).

(10) Update of Item History and Performance

Each time an item is selected for use, information regarding the date of administration, client choosing the item, group to be tested, and item statistics may be recorded. This information is valuable in maintaining the bank, providing data concerning changes in testing emphasis, detecting changes in student performance, monitoring item use, and identifying previously undetected item flaws. Baker (1974) has suggested a number of statistics that may be recorded in an item history.

(11) Periodic Review and Modification of Item Bank

This step is included to stress the importance of regular review of the items in the bank and of the services offered by the bank. This review will make it possible to identify areas for which more items are needed, anticipation of new areas in which items are to be developed, and plans for additional services to be incorporated into the bank.

Item banking is becoming increasingly popular with test developers and users alike (Hambleton and Cook, 1977; Hiscox and Brezinski, 1980; Rentz and Beshaw, 1977). There are several reasons for the popularity of item banks. One advantage for test developers is that the need for hiring and training item writers, conducting pilot studies, and validating item content for new tests can be reduced. During the test development process, many more items are usually written than are actually needed. Pilot administrations are used to identify items which are of high enough quality to be included in the final test and to identify flaws in items. Items which are not used in

test development, but which are of adequate quality can be placed in an item bank. Other items, after revision and repiloting, can also be placed in the item bank. When test revision or development of additional test forms is being considered, the items in the bank can be used without need for a subsequent pilot administration.

The concept of an item bank is an attractive one. There is great appeal to the notion of having a large collection of high quality items for use in building or revising tests. Each time an item is used, more information is gained about strengths and possible flaws in the item. Statistical information can be updated and monitored with each administration. After an item is used, it can be returned to the bank and replaced with a similar item. It would seem that anyone involved in test construction would develop an item bank. Test developers have, in fact, developed large collections of test items and stored them in such a way as to facilitate test construction and modification (Henrysson, 1971). Such item collections have made it possible for test developers to monitor item use and organize items by content for revision. Unfortunately, statistical problems have prohibited item banks from achieving their full potential.

Statistics of item difficulty and discrimination which are traditionally used to describe items are dependent on the sample of examinees who attempt the item. Item difficulty, for example, is generally denoted by "p," the percent of examinees who successfully attempt an item. An item will have a higher p-value in a more able group of examinees than in a group of less able ones. If different items are pilot tested with different samples of examinees, the p-values are not comparable. Although classical test theory has allowed us to

make comparisons of examinees who take different tests (through various equating techniques) similar techniques for comparing items are not available with classical methods. Thus, integrating items into an item collection is not totally satisfactory. Through pilot testing an indication of item quality is available, but the relationship of the item difficulty of new items to those already in the item collection or on other test forms is not known. Although items can be piloted with several diverse groups, the results are not satisfactory (Choppin, 1976). Item discrimination, which is usually taken as the correlation between performance on an item and total test performance, has the additional disadvantage of being dependent on the other items in the test.

Partially as a result of these statistical problems, item banking has never progressed far beyond the stage of being more than a collection of items available for use. When items are pulled from the collection to construct a new test or to replace items in a current test, little has been known about the effect of the change. Thus, until after the new or revised test has been administered, no comparisons between the new and old tests can be made. Without sample independent item statistics (item statistics that do not depend on the sample of examinees who attempt the item or on the other items in the test), the items in the bank are not comparable.

Sample independent item statistics are among the theoretical advantages offered by a relatively new area of measurement research, item response theory (also called item characteristic curve or latent trait theory). During the last ten to twelve years, item response theory has developed from a topic known only to a few measurement

specialists to a widely applied practical tool (Hambleton, 1979). While much research and developmental work is still being performed, the technology is sufficiently developed for the use of item response theory in many applications (Hambleton, 1980; Lord, 1980).

Item response theory assumes that examinee performance on a test is explained by an unobservable characteristic or trait. The item response models that are currently well enough developed for practical use are unidimensional; that is, they assume only one trait underlies test performance. Test performance is an observable demonstration of an examinee's unobservable ability or trait. An item response model specifies a mathematical relationship between the observable test performance and the unobservable trait. The exact form of the mathematical relationship is determined by the particular model chosen to represent the data. The most familiar and widely used models are the one-, two-, and three-parameter logistic models. (The one-parameter model is also called the Rasch model.)

Several advantages of item response theory make it an attractive alternative to classical test theory. Two of the advantages are (1) the statistics used to describe a test item are independent of the examinee group that attempted the item, and (2) the estimate of an examinee's ability is independent of the particular test items he or she attempted. Thus, item statistics from two different tests of the same trait can, after calibration, be compared. Similarly, examinees who attempt different sets of items can be compared if the items have been calibrated to the same scale. The extent to which these advantages are obtained in practice depends on the extent to which the model being used fits the data. If there is a poor fit between a model and

the data, the advantages of item response theory will not be obtained or will be poorly attained. Introductions to item response theory, assumptions, models, and advantages over classical test theory are available from numerous sources. See, for example, Allen and Yen (1979), Hambleton (1979), Hambleton and Cook (1977), Lord and Novick (1968), Wright (1968, 1977), and Wright and Stone (1979). In the next section of this paper, the use of item response theory in calibrating an item bank is discussed.

Calibration of Items in an Item Bank

As was explained above, for an item bank to be maximally useful, the statistics which describe the items must be comparable from item to item. That is, the statistics must be on the same scale. When items are analyzed together, from the results of a single pilot with a single examinee group, the results are comparable. More often, however, a pilot administration involves several groups of examinees taking different sets of items. In the case where items are to be integrated into an already existing item bank, the pilot administration will almost certainly involve examinees different from those used to initially stock the bank. Therefore, a method is needed to calibrate the item statistics to a single scale.

A common design for pilot administrations involves having different examinee groups attempt a block of common items and a block of items which are not common to the two tests. Thus, the common items serve as a "link" between the other items. If there are more than two examinee groups, it is possible to have one block of items

common to all groups or to have several linking groups of items. In the latter case each examinee group is linked to only two other groups by common items.

The method of calibrating items to a single scale using item response theory when two groups of examinees have taken some common and some different items is described in Cook (1981). Suppose two groups of examinees, X and Y, have taken pilot tests. The two pilot tests contain some common and some different items. First, the following values are calculated for the common items:

$$M_{b_x},\ SD_{b_x},\ M_{b_y},\ \text{and}\ SD_{b_y}.$$

Here, M and SD represent means and standard deviation; X and Y are the examinee groups, and b represents item difficulty. Next, to convert item statistics (b=difficulty, a=discrimination, c=pseudochance or guessing) from the scale of X to the scale of Y, the following formulas may be used:

$$b_y = Ab_x + B \qquad (1)$$

where $A = SD_{b_y}/SD_{b_x}$.

and $B = M_{b_y} - AM_{b_x}$.

$$a_y = a_x\ (1/A) \qquad (2)$$

$$c_y = c_x \qquad (3)$$

Ability estimates (θ) are adjusted as follows:

$$\theta_y = A\theta_x + B, \qquad (4)$$

In the next section the use of item response theory in the development and use of the Louisiana Item Bank is explained.

Development and Use of the Louisiana Item Bank

In April of 1981 the Louisiana Department of Education conducted an item pilot of second grade test items. A total of 422 items were piloted in twelve test forms. Six forms (A-F) contained language arts items (162 items) and six (G-L) contained mathematics items (260 items). A twenty percent random sample of Louisiana second graders was selected to participate in the pilot. A total of 11,055 usable student records were obtained from the administration of the pilot tests. The number of students attempting each test form, the items included in the form, and test statistics can be found in Table 1. Analyses of language arts and mathematics were conducted separately.

Following classical item analysis, item response analysis was conducted. The LOGIST computer program (Wood, Wingersky, and Lord, 1976) was used to obtain item parameter estimates for all items. Once item parameter estimates were available, several measures of goodness of fit were calculated. One of these measures involved plotting a predicted item characteristic curve against the obtained curve. To do this, the ability scale was divided into equal intervals. A predicted proportion correct (P) was calculated for the ability (O) at the midpoint of each interval using the formula:

$$P = c + (1.0 - c)\ ((e^{Da(\theta - b)}) / (1 + e^{Da(\theta - b)})) \quad (5)$$

where D, the only term not previously defined is 1.7. This value was calculated for every interval and every item. The observed proportion correct was calculated for each interval by determining the proportion correct on each item for examinees with ability estimates in that interval. These values were plotted against each other. Figures 4, 5, and 6 are examples of some of the resulting plots. In the Figures, the symbol "*" denotes predicted values while a "+" is an observed value. Observed values that were not close to predicted values were taken as indications of misfit and, if they occurred often for an item or affected a large proportion of the examinees, served as a basis for subjecting the item to further review and possible revision and re-piloting. For the majority of the items, however, fit was judged to be adequate. Following goodness of fit assessment, items were calibrated to a single scale in each subject area. The calibration occurred as outlined in the previous section of the paper.

Following calibration of the items, selection of items for the test occurred. From the pool of items measuring each objective, four items were chosen. Three issues were considered in item selection. These were content representativeness, classical statistics, and item response analysis. Although items were written in accordance with detailed item specifications, it was sometimes apparent that a particular item was not representative of the other items for that objective. Such items were not included in the test. Such items were identified mainly by (1) difficulty values (p-values) which were much different from the difficulties of other items measuring the objective, (2) poor fit between

Table 1
Test Statistics, Number of Examinees, and Items
In the Second Grade Pilot Tests

Test Form	Number of Examinees	Total Items	Items in Form	Mean	SD	KR-20	SEM
A	880	30	1-10, 21-30	26.86	5.454	.940	1.856
B	890	30	1-10, 31-50	27.26	5.114	.940	1.740
C	949	42	1-10, 51-82	37.53	6.591	.936	2.312
D	876	35	1-10, 83-107	28.78	8.013	.951	2.481
E	888	35	1-10, 108-132	30.89	6.467	.943	2.146
F	959	40	1-10, 133-162	34.71	8.458	.962	2.323
G	961	60	163-172, 172-222	51.83	9.324	.948	2.959
H	961	50	163-172, 223-262	44.70	7.387	.936	2.597
I	938	50	163-172, 263-302	41.28	8.494	.938	2.947
J	889	49	163-172, 303-342	44.98	7.289	.944	2.406
K	943	50	163-172, 343-382	42.98	7.295	.924	2.782
L	921	50	163-172, 383,422	43.29	8.423	.941	2.848

[1] This form actually contained 50 items, but Item 171 was deleted from the analysis due to a scanning error.

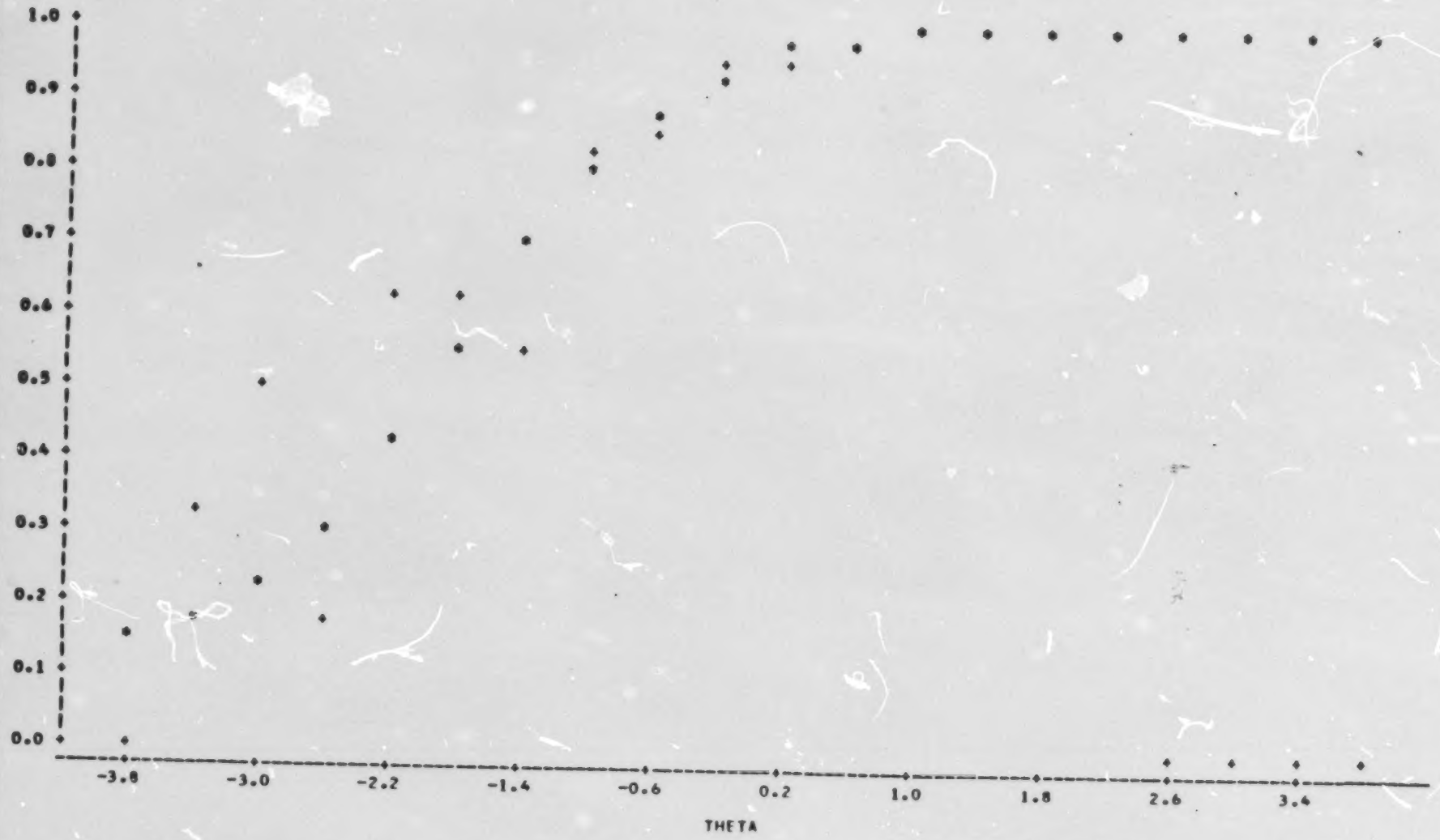


Figure 4: Goodness of Fit Scatterplot - Item 165

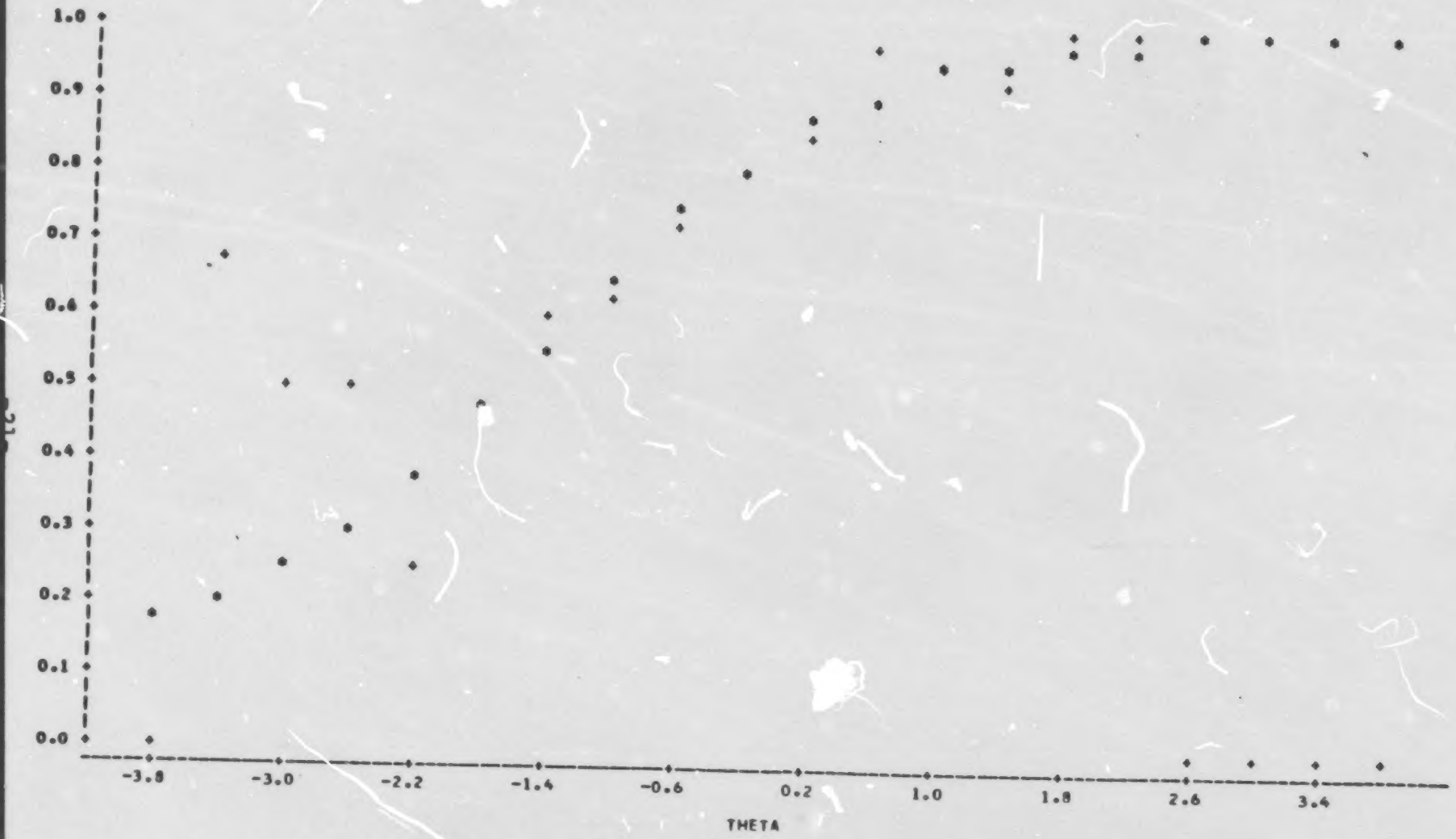


Figure 5: Goodness of Fit Scatterplot - Item 166

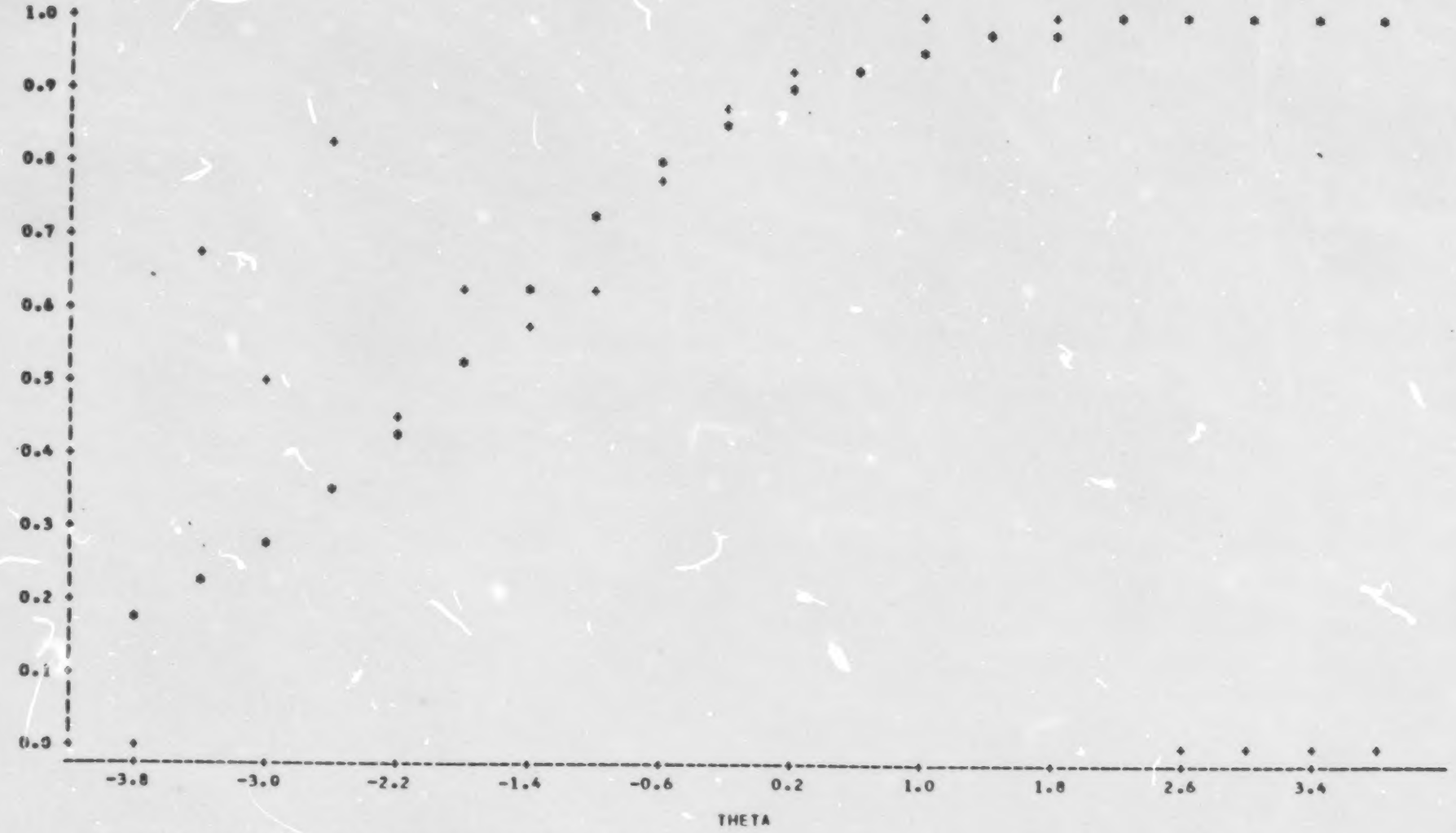


Figure 6: Goodness of Fit Scatterplot - Item 171

the data and the item response model, and/or (3) logical analysis of the item content and options. The final second grade test consists of two 60-item sections.

For each section of the test, language arts and mathematics, it was necessary to establish a performance standard. It was necessary to establish the standard since the tests are to serve as a part of the promotion decision and to target students for remediation. One of the first uses of the calibrated item bank was in establishing the performance standard.

The final grade two test is composed of some items from every test form that was piloted. Therefore, no student took all the items. One of the advantages, however, of item response theory is that an examinee's ability estimate is independent of the specific items from a calibrated pool that he or she attempted. Thus, performance on one set of items in an item pool can be inferred from observed performance on another set. This advantage was used in the determination of the performance standard.

For every examinee in the pilot sample an ability estimate (θ) was available for the subject area in which a pilot test was taken. Also available were item parameter estimates (b, a, and c) for each item selected for the final test. Using Equation 5 it was possible, therefore, to predict the probability of each examinee answering each item correctly. By summing these probabilities for all sixty items in the test, it was possible to obtain a predicted test score for each examinee. Also available was a prediction of each student's test performance as clearly passing, clearly not passing, or borderline. This prediction was made by each student's classroom teacher. Thus,

it was possible to predict a test score distribution for each of the three performance groups. By comparing the predictions of score distributions for clear passers and non-passers, decisions regarding the performance standards were made (Mills, 1981).

Item response theory will also play a part in the revision of the basic skills tests. Each year the tests will be revised using items in the bank. Since the item parameters of these items are already known, it will be possible to predict whether or not the revised tests are easier or harder than the current ones. If differences in test difficulty are predicted, adjustments can be made to the performance standard. Another future use of item response theory in the item bank will be in the addition of new items to the bank. Even though more items were piloted than are needed in any one year, there will be a need to pilot more items and add them to the bank. There are two reasons for this. First, some of the items included in the pilot were judged to be in need of revision and re-piloting before use. Second, since the tests will be revised annually, it will not be too many years before all items have been used. Of course, items can be re-used after an idle period of a year or two, but for some objectives more items are needed to allow for maximum flexibility in test revision.

Summary

During the last ten to twelve years much psychometric research has focused on the topic of item response theory. Developments in this area have made possible advances in the use of item banks (and in many other areas). Two important advantages of item response theory are that estimates of examinee ability are not dependent on the specific

items attempted, and item statistics are independent of the examinee group. The Louisiana Department of Education has used item response theory in the calibration of an item bank for use with its Basic Skills Testing Program. Item response theory was used in calibration of the items to a single scale in each subject area and for modeling test performance to assist in determining a performance standard. Future applications of item response theory with the bank will be in test revision and integration of new items into the bank.

References

Allen, M. J. & Yen, Y. M. Introduction to measurement theory. Monterey, CA: Brooks/Cole Publishing Company, 1979.

Baker, F. B. The roles of statistics, In G. Lippey (Ed.), Computer-assisted test construction. Englewood Cliffs, NJ: Educational Technology Publications, 1974.

Choppin, B. H. Recent developments in item banking: A review. In D. N. M. De Gruijter and L. J. T. van der Kamp (Eds.), Advances in psychological and educational measurement. London: John Wiley and Sons, 1976.

Cook, L. The application of item response theory to test equating Presentation in the training session: Item response theory: Concepts tools, and applications. A presession at the annual meeting of the American Educational Research Association, Los Angeles, 1981.

Hiscox, M. & Brezenski, E. A guide to item banking. Portland, Oregon: Northwest Regional Educational Laboratory, 1980.

Hambleton, R. K. Latent trait models and their applications. In R. Traub (Ed.). New directions for testing and measurement: Methological developments. San Francisco: Jossey-Bass, 1979.

Hambleton, R. K. Latent ability scales, interpretations, and uses. In S. Mayo (Ed.), New directions for testing and measurement: Interpreting test performance. San Francisco: Jossey-Bass, 1980.

Hambleton, R. K. & Cook, L. Latent trait models and their uses in the analysis of educational test data. Journal of Educational Measurement. 1977, 14, 75-96.

Hambleton, R. K. & Eignor, D. R. A practitioner's guide to criterion-referenced test development, validation, and test score usage. Laboratory of Psychometric and Evaluative Research Report No. 70. Amherst, MA: School of Education, University of Massachusetts, 1979.

Henrysson, S. Gathering, analyzing and using data on test items. In R. L. Thorndike (Ed.) Educational Measurement Washington, D.C.: American Council on Education, 1971.

Lord, F. M. Applications of item response theory to practical testing problems. New York: Lawrence Erlebaum Associates, Inc., 1980.

Lord, F. M. & Novick, M. R. Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

Mills, C. N. Determination of performance standards for the second grade basic skills tests. Bureau of Accountability Technical Report No. 1. Baton Rouge, LA: Louisiana Department of Education, 1981.

Popham, W. J. Criterion-referenced measurement. Englewood Cliffs, NJ: Prentice-Hall, 1978.

Prosser, F. Item banking. In G. Lippey (Ed.), Computer-assisted test construction. Englewood Cliffs, NJ: Educational Technology Publications, 1974.

Rentz, R. R. & Bashaw, W. L. Equating reading tests with the Rasch model. Athens, GA: Educational Research Laboratory, 1975.

Shoemaker, D. M. Applicability of item banking and matrix sampling to educational assessments. In D. N. M. De Gruijter and L. J. T. van der Kamp (Eds.). Advances in psychological and educational measurement. London: John Wiley and Sons, 1976.

Stodola, O. C. Item classification and selection. In G. Lippey (Ed.), Computer-assisted test construction. Englewood Cliffs, NJ: Educational Technology Publications, 1974.

Thorndike, R. L. Educational measurement for the seventies. In R. L. Thorndike (Ed.), Educational measurement, Washington, D.C.: American Council on Education, 1971.

Tinkelman, S. N. Planning the objective test. In R. L. Thorndike (Ed.). Educational measurement. Washington, D.C.: American Council on Education, 1971.

Wood, R. L., Wingersky, M. S. & Lord, F. M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters. Research memorandum 76-6. Princeton, NJ: Educational Testing Service, 1976.

Wright, B. D. Sample-free test calibration and person measurement. In Proceedings of the 1976 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service, 1968.

Wright, B. D. Solving measurement problems with the Rasch model, Journal of Educational Measurement, 1977, 14, 97-116.

Wright, B. D. & Stone, M. H. Best test design. Chicago: MESA Press, 1979.

THE EFFECTS OF PROMOTIONAL TESTING ON A LARGE CITY SCHOOL SYSTEM

By

Ellen Pechman

New Orleans Public Schools

Paper presented at the Annual Meeting of the American Educational Research Association, New York, 1982 (Session 10.27). The opinions presented here are solely the author's and should not be interpreted to represent those of the Orleans Parish School Board.

I am pleased to have the opportunity to participate with my State Department colleagues in this session to discuss the local effects of Louisiana's recently implemented promotional testing program. I think the program has both practical and conceptual considerations of real importance, and my discussion will touch on both kinds of implications. I will begin with a little history of accountability activity in New Orleans.

The comprehensive accountability actions mandated by Act 750 -- the 1979 enabling legislation which called for the promotional testing program required New Orleans to make relatively few changes in its policies or practices. We established a norm-referenced testing program for all students in the district in 1975; and for the past five years, these test results have been used to guide instructional planning for individual children and for the system as a whole. There have also been Board-defined promotional policies in the district for many years, and these policies were being regularly revised and updated well before the recent period of concern about accountability. For several decades, we have funded and conducted numerous types of remedial activities for children who were having difficulties in school. Many of the programs have been paid for with local funds and, of course, we have been the beneficiaries of sizeable federal support for remediation which has been very helpful. Finally, since the late 1970's we have also been strengthening the basic or minimum skills aspects of our instruction, and our efforts in the curriculum area coincided with the legislative activity that led to Act 750. Thus, as the law has been phased in, it has reinforced and supported already existing curriculum and policy developments initiated in our district. Still the promotional testing aspect of Act 750 has impacted us in some important ways.

The effect of Act 750 was first felt in New Orleans, as in other districts across the state, when the State directed us to re-examine and redefine our existing promotional policies according to a specified procedure which included substantial input and involvement of the parent community. Thus, we established a committee which included representatives from various major organizations and constituencies, and we spent several months reviewing policies and, where necessary, adjusting our requirements to coincide with those mandated by the State. Although it is always a little cumbersome to make plans and decisions within the structure of a large committee, we had excellent support from parent and community groups and our policy revisions were successfully made according to the new requirements.

On the whole, we felt the revision process was very positive, although the only real changes we made were not based upon community input or on committee opinion, but, instead, on the State regulations themselves. Specifically, two substantive changes were made in the promotion rules: 1) in the curriculum area, mathematics and language arts were added to reading as promotional subjects, and 2) the state tests were described as being the "principal criteria" for promotion. What are the implications of these few changes? I think there are several of key importance.

First, the addition of mathematics as a promotional subject for second graders might, I think, be the most fundamental and substantive change made. Before this year, reading was the major subject area on which second grade students were evaluated for promotion. The non-math promotion policy grew out of research in recent years that recorded the delay in children's concept readiness for math. Because developmental psychologists and math researchers were telling us that children didn't usually attain the logical framework for

math learning until between the ages of six and eight, it made sense to hold off requiring math for promotion until grade three. This would, at the same time, enable us to make reading the major focus of the basic skills development during the first several years of school. Although mathematics has been included in the curriculum beginning in kindergarten, our instructional strategies and policies were oriented towards developmental and readiness activities. When children were identified as slow to respond to the math readiness programs, remediation was provided, even in the first grade, but it was not a promotional consideration. Now, as a result of the new state requirements, the expectation for children's developmental maturity in mathematics has been accelerated and has become an integral part of the decision regarding student's promotion to grade three.

While some are applauding this move as a return to the socalled "basics," I fear the change may not necessarily have the expected positive results. According to our own standardized testing, about 35 40 percent of our students do not have the readiness skills for grade level work in mathematics when they enter school. This, of course, is because so many of our students do not get the extensive learning foundation at home that some families can give their youngsters before the school years begin. Thus, I am concerned that this change in the requirements will cause our second graders to be drilled on the 15 mathematics and 15 language arts objectives so they can pass the test. Unfortunately, the skills learned through isolated drills are not likely to be long lasting. Recent findings from the National Assessment raises fundamental questions about whether an exclusively skills-focused curriculum in the early years provides the needed grounding in context learning and conceptual analysis for the skills to be productively applied as children grow older.

The new requirement that the tests become the "principal criteria" for promotion is the second major change incorporated into our promotional rules. While, in one sense, this is a very minor portion of our overall promotion plan, in another it is probably the most dramatic element. The promotional test has been the part of the total accountability package that has captured the greatest interest and caused most alarm among teachers and the general public. Act 750 explicitly keeps the final decision regarding promotion in teachers' hands, stating that "each teacher shall, on an individual basis, determine the promotion or placement of each student," allowing only for review by a school-level representatative promotional committee. However, the State Department of Education established implementation rules which required local school systems to include the tests as the "principal criteria" for promotion in their written policies. While this has in fact meant that the tests are included as one of the several criteria for promotion, including classroom performance, attendance, other test information and so on, the wording of the requirement has caused confusion and tremendous concern in the school community.

While basic skills testing is a fundamental component of good instructional practice, linking such testing to promotion unfortunately lends itself to easier sloganeering than to any real improvement in teaching or learning, and the test has been attributed far more power than it has. As educators, we know that to base a child's promotion on a single test is entirely inappropriate, but there are many people -- some educators and certainly many parents and members of the community -- who believe that a test can be so used and, in this way, solves the "social promotion" problem. This claim is questionable for all children, but particularly so in regard to young children who do not have the cognitive skills that enable them to think about and control their own studying habits and thinking capacities.

My hope is that, somehow, in the implementation of the testing program, we will be able to keep our perspective about the limitations of tests that I have mentioned, and maintain the potential usefulness of conducting an annual basic skills assessment in kindergarten through the twelth grade. But, for the time being, I am concerned that we have become too absorbed with the promotional aspects of the Basic Skills Test. The publicity accompanying the testing program has inflated the test itself out of proportion to its actual importance, and, unfortunately, has caused the 30 testable objectives to become strongly emphasized in classrooms across the district. We are hard pressed to keep a reasonable calm about the test, especially with the political interest and pressure associated with it. All the administrative directives and supervisory platitudes about curriculum will have difficulty countering the intimidating public relations that has accompanied the state test. Earlier this fall, our own State Superintendent of Education contributed to the confusion by coming into the district and reporting at a press conference that, according to his staff's predictions, about 30% of New Orleans students would fail the state test. This got picked up in the newspapers as "30% of Orleans second graders will flunk second grade." Having worked closely with Mr. Nix's staff, I knew no such predictions had been made, but my corrections of these "facts" were issued to deaf ears.

My concern is that the legislative objective of Act 750 -- to strengthen educational programs -- may backfire on us. Focus on promotional testing may cause the comprehensive curriculum to give way to a thirty objective curriculum. If such a situation occurs, public schools will be providing a dull and unnourishing educational diet. Undue emphasis on testable objectives, taught in isolation from the substantive context, will make the curriculum too narrow for able students and even more difficult to comprehend for the weaker students who we are trying most to help. Although no formal changes have been made in the New

Orleans curriculum and teaching plan, I am concerned about the informal adjustments that have been made.

At the administrative level we have worked hard to stress the importance of continuing to teach the complete curriculum and to avoid the temptation to teach only the test. One of the major positive changes brought by Act 750 was that curriculum revisions were mandated. The State Department of Education wrote and disseminated explicitly stated curriculum standards and new curriculum guides which include suggested teaching activities and instructional materials that were sent into the parishes. We now have clear and concise lists of the minimum skills that form the foundation of a comprehensive statewide instructional program from Kindergarten through Grade 12. This effort coincides with our local initiatives and goals in the curricular area and, at the district level, we are emphasizing the importance of the minimum skills as a foundation of a wide-reaching curriculum.

Remediation is a third dimension of the pupil progression plan that has been influenced by the regulations accompanying the implementation of the state testing program. The legislation provides funds to support remediation for students who "fail to meet the minimum mastery levels." Since the problem of student achievement is one we in New Orleans have faced for many years, we already have numerous developmental and remedial programs designed for "at risk" students, beginning as early as Grade 1. The state funds accompanying the testing program are earmarked for use only on students who fail the promotional test, thereby requiring the district to establish a supplemental remedial effort which stands apart from the continuing effort in this direction. Since we won't know who fails the test until after the end of the year, the remediation will be offered in the summer, well after the failure has occurred. Our preference would be to use state funds to provide remediation for students before the failure and accompanying labeling occurs.

While additional funds for remediation are always appreciated in a district such as ours, certain logistical and curricular problems are also posed by this type of supplemental funding. Since state funds can only be used for students who fail the state test, we must establish special summer programs for this group. If students are placed apart in state-funded remedial programs, it is logical that teachers will emphasize the nonmastered objectives, and not necessarily the broadbased diagnostic needs of children. Such an objective-centered approach will be of greatest assistance to borderline achievers who are just shy of meeting minimum requirements and need only a little additional instructional assistance to bring their performance to grade level. But this type of program is unlikely to help the students who have more comprehensive achievement problems and who missed the 75% mastery target by a great deal. Thus, it is not clear if summer school remediation can sufficiently assist the students who are most in need of educational support.

An additional important issue which is currently concerning us is the planning of the summer programs. The funds, of course, will not be available until after the results of the test are in, so it is quite a challenge to initiate the kind of comprehensive planning needed to undertake a smooth running program in the summer. We are making very general estimates of the size and locations of our programs, but it will be enormously difficult for us to have classes staffed, supplied and implemented, especially in a district of our size, in the six weeks followirg the date we receive our test results and identify the students who require remediation. Remediation planning efforts are well underway, but we are groping in the dark and it is difficult to determine how effectively or successfully we will be able to use the resources provided for the targeted students.

Finally, I want to speak to the problems of administration that accompany the implementation of a large scale testing program, especially one which is as sensitive and as highly charged as this one. Issues relating to test security,

consistency and standardization of test administration, monitoring, and even the simple logistics associated with packaging and assuring the accuracy of the computer scanned test responses are all major concerns. Any one of these areas can become problematic even in a small school district, and the problems are multiplied each time an additional test administrator, test coordinator or principal is interpreting the procedures.

Of course, the expected questions have surfaced regarding the rules for testing special education and limited English speaking populations. While, as you have heard, a good deal of effort has been put into clarifying these issues at the state level, at the classroom administration level we depend upon teachers' understanding of what they read. Even with experienced and specially trained test administrators, consistency of test implementation is hard to achieve. Classroom teachers with little testing experience will be expected to intuitively follow the twists and turns of test procedure and to precisely administer the tests according to all the specifications of the written rules. Even the clearest writing effort in testing produces a language foreign to many teachers, and when the rules are applied with groups of several hundred children in schools, especially with young children, we can't be entirely certain of the outcome.

Under the best of circumstances, the implementation of a testing program across a school system our size -- we have almost 7000 second graders -- is a major effort. In a situation such as this one, every small issue is greatly magnified in importance. Added central office and school level personnel are needed to coordinate and manage the flow of information and materials. While central support has been slightly augmented at the school level, additional staff support and release time are required if the most careful implementation of the test procedures is to be assured. But neither additional staffing nor

release time has been funded by Act 750, so we will lean on the generosity, dedication and good intuition of a few key teachers and support personnel at each school. In the central office, we have already spent several months of staff time and thousands of dollars producing curriculum guides and instructional materials. Soon more money will be invested to check, sort and package test materials when they arrive in the district. In schools, substantial effort must be made to recheck children's computer scanned responses to assure that the coded information is clearly machine readable so results will accurately reflect what students intended to write. No dollar figure has been projected for the cost of this effort in staff time and, particularly, for teaching time lost to school level planning and logistical activities.

Some will argue that the costs of this testing program are outweighed by the benefits, especially because of the diagnostic information the test will provide. I hope this is true, but I am skeptical. If we make the assumption the test has the psychometric quality we expect and that, indeed, the testable objectives represent the central instructional core of knowledge children need for later learning assumptions which remain to be validated -- still, the tests only provide us with information which every good classroom teacher routinely collects during the school year. By May, when we receive the diagnostic feedback, the data will be long past its usefulness. Teachers already well know whether or not their students can read and do mathematics, and the substantive information on which to make the promotional decisions has already been reported to parents. The test will merely confirm the label of failure on those who have been failing all along, and it will mean little to those who are succeeding.

In a large school district which is short of funds, high in needs, and struggling to meet complex and conflicting human and instructional demands, a testing program is an emotionally and financially draining experience. Testing

is nonetheless essential and must be conducted if we are to have the objective information necessary for planning comprehensive instruction. The question must be asked, however, if a promotion-focused testing program gives a good return for the effort. "We are full of the basics," argues a colleague of mine, "it's the complexities which are sorely lacking." If we are indeed neglecting the complex issues regarding the education of children, is the focus on a promotional testing program working on behalf of educational improvement or against it? This is the question which must be kept in mind as we observe the implementation of this program. It will be particularly important for us to be honestly alert to the program impact and to be wise enough to make adjustments as they become necessary and before such changes are too late.